Genetica **112-113:** 199–222, 2001.

Explaining stasis: microevolutionary studies in natural populations

J. Merilä[1]*, B.C. Sheldon[2] & L.E.B. Kruuk[3]
[1]*Department of Population Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18 d, SE-752 36 Uppsala, Sweden;* [2]*Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK;* [3]*Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, UK;* **Current address: Department of Ecology and Systematics, Division of Population Biology, P.O. Box 17 (Arkadiankatu 7), FIN-00014 University of Helsinki, Finland (Fax: +358-9-191 28701; E-mail: juha.merila@helsinki.fi)*

Key words: animal model, breeding value, *Cervus elaphus*, *Ficedula albicollis*, microevolution, natural selection

Abstract

Microevolution, defined as a change in the genetic constitution of a population over time, is considered to be of commonplace occurrence in nature. Its ubiquity can be inferred from the observation that quantitative genetic divergence among populations usually exceeds that to be expected due to genetic drift alone, and from numerous observations and experiments consistent with local adaptation. Experimental manipulations in natural populations have provided evidence that rapid evolutionary responses may occur in the wild. However, there are remarkably few cases where direct observations of natural populations have revealed microevolutionary changes occurring, despite the frequent demonstration of additive genetic variation and strong directional selection for particular traits. Those few cases where responses congruent with expectation have been demonstrated are restricted to changes over one generation. In this article we focus on possible explanations as to why heritable traits under apparently strong directional selection often fail to show the expected evolutionary response. To date, few of these explanations for apparent stasis have been amenable to empirical testing. We describe new methods, derived from procedures developed by animal breeding scientists, which can be used to address these explanations, and illustrate the approach with examples from long-term studies of collared flycatchers (*Ficedula albicollis*) and red deer (*Cervus elaphus*). Understanding why most intensively studied natural populations do not appear to be evolving is an important challenge for evolutionary biology.

Introduction

One of the main goals of evolutionary biology is to understand how the interplay between natural selection and inheritance translates into evolutionary change and thus contributes to adaptation. In a simplified quantitative genetic framework, this reduces to understanding the ecological factors responsible for determining the intensity of natural selection acting on traits, as well the processes and factors influencing the amount of heritable genetic variation in the trait. The evolutionary response (R) over one generation in a given trait is predicted to be (Falconer & Mackay, 1996):

$$R = h^2 S, \tag{1}$$

where h^2 represents the heritability of the trait, and S is the directional selection differential. If natural selection acts consistently on a heritable trait, one should thus expect the mean value of the trait to change over time. However, this simple prediction is further complicated by the fact that different traits can be genetically correlated, and the expected response to selection is therefore better described by a multivariate extension of Equation (1) (Lynch & Walsh, 1998):

$$\Delta \mathbf{z} = \mathbf{G}\boldsymbol{\beta}, \tag{2}$$

where $\Delta \mathbf{z}$ is the vector of expected selection responses, $\mathbf{G}$ is the additive genetic variance-covariance matrix, and $\boldsymbol{\beta}$ is the vector of selection gradients. By

expanding this expression we get (Lynch & Walsh, 1998):

$$\begin{bmatrix} \Delta z_1 \\ \Delta z_2 \end{bmatrix} = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \quad (3)$$

where Δz_1 and Δz_2 are the predicted responses in traits 1 and 2 respectively, G_{11} and G_{22} are the corresponding additive genetic variances, G_{12} and G_{21} refer to the additive genetic covariance among these two traits, and β_1 and β_2 are the selection gradients (as estimated from the multiple regression of trait values on fitness (Lynch & Walsh, 1998)) for these two traits. From this equation, it is obvious that the expected response in a given character depends not only on the amount of additive variance and the force of selection acting on it, but also on the sign and magnitude of genetic covariance between traits, as well as on the force of selection acting on correlated traits.

In spite of the fact that these quantitative genetic models (Eqs. (1) and (2)) of evolutionary change have considerable support from laboratory studies (e.g., Falconer & Mackay, 1996; Lynch & Walsh, 1998), direct evidence for microevolutionary change in natural populations obeying these simple equations is limited (Grant & Grant, 1993, 1995; see also Hairston & Walton, 1986). However, a large body of evidence suggests, indirectly, that microevolutionary changes over short time periods are of commonplace occurrence. For example, there is widespread evidence for local adaptive genetic differentiation among contemporary populations of various organisms (e.g., Berven & Gill, 1983; Berthold et al., 1992; Ebert, 1995; De Meester, 1996; Linhart & Grant, 1996; Lively & Jokela, 1996; Bone & Farres, 2001; Reznick & Ghalambor, 2001), although the time-scale underlying this differentiation is often unknown (see Hendry & Kinnison, 1999; Huey et al., 2000; Bone & Farres, 2001 for exceptions). Furthermore, a compilation of data from studies which have compared the degree of genetic differentiation in ecologically important (quantitative) traits among different populations of the same species show that this differentiation normally exceeds that to be expected due to genetic drift alone (i.e., the degree of differentiation in neutral marker loci, Figure 1). Consequently, unless there is a publication bias towards studies finding high divergence between populations, local adaptation, and hence microevolution, is apparently common. Another interesting feature of these data (Figure 1) is that, perhaps in contrast to current consensus (e.g., Butlin & Tregenza, 1998; Hedrick, 1999), the degree of divergence

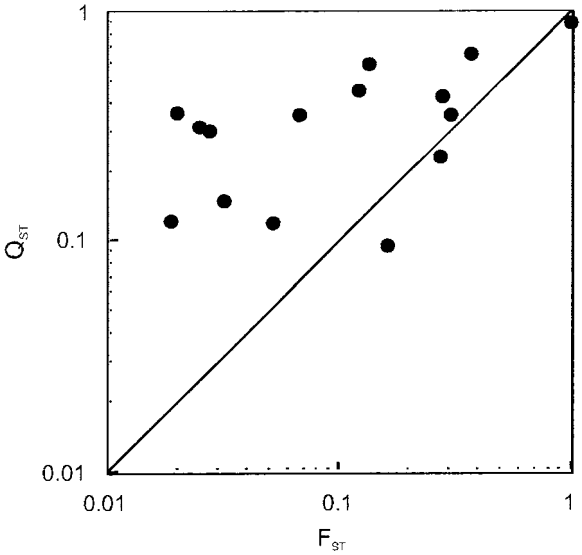

Figure 1. Comparison of quantitative trait (Q_{ST}) and molecular (F_{ST}) genetic differentiation among local populations of the same species. F_{ST} gives the expected degree of differentiation due to genetic drift alone above which natural selection needs to be invoked to explain the differentiation (e.g., Spitze, 1993). Each point represents one species, and the solid line indicates 1:1 correspondence. Although the Q_{ST} values are on average higher than F_{ST} values (Wilcoxon signed-rank $z = 2.63$, $p = 0.0086$), differentiation in the two trait classes is highly correlated across the studies ($r_s = 0.71$, $p = 0.0136$). Data adapted from Lynch et al. (1999), with additional data from Merilä (1997), Waldmann and Andersson (1998) and Andersson, Waldmann and Pretince (2000).

in single locus molecular genetic markers predicts the degree of divergence in quantitative traits $r_s = 0.71$, $z = 2.474$, $p = 0.0136$). By inference, since molecular data suggest that the degree of genetic differentiation among many contemporary populations is moderate to large (e.g., Ward, Skibinsky & Woodwark, 1992), natural selection promoting genetic differentiation in quantitative traits can also be expected to be common.

A number of recent experimental studies suggests that microevolutionary responses in natural populations may be rapid. These studies have generally taken the form of observing the response of populations either deliberately (Losos, Warheit & Schoener, 1997; Reznick et al., 1997) or inadvertently (Huey et al., 2000) introduced to new environments. Although these studies are important, in that they can show both the potential rapidity of evolution in the wild, and establish that quantitative genetic change has occurred in those traits which have shown phenotypic responses (Huey et al., 2000), they generally can not identify the causal selective factors responsible for the change, particularly if replication is lacking

(but see McKenzie & Batterhamn, 1994). Nor do they allow a test of the predictive power of quantitative genetic theory, because they do not measure selection gradients, but instead infer selection intensities from the measured phenotypic response in combination with measurement of additive genetic variances (e.g., Reznick et al., 1997).

Endler (1986) showed that natural selection was common in natural populations. A more recent compilation (Kingsolver et al., 2001; see also Kinnison & Hendry, 2001) reinforces this view, although it also suggests that strong natural selection may not be particularly common since the median standardized selection intensity (i), based on more than 2500 estimates, was only 0.13. Selection of this strength, coupled with a heritability typical for a morphological character in a wild population (≈ 0.4) is expected to lead to an evolutionary response of ≈ 0.05 phenotypic standard deviations per generation (or 'haldanes', Hendry & Kinnison, 1999). Hendry and Kinnison's (1999) review of the rates of microevolution in contemporary populations documented very few cases of evolution at this rate, except for those restricted to a very small number of generations (e.g., Endler, 1980; Grant & Grant, 1995). In this paper we adopt the median strength of directional selection identified by Kingsolver et al. (2001) as a division between strong and weak natural selection in natural populations. The main point of our treatment is that there are a large number of traits in natural populations which appear to be subject to strong natural selection, and to possess abundant additive genetic variation, but which do not seem to be evolving. This is a familiar paradox, and there are several potential explanations for the observed stasis (e.g., Price, Kirkpatrick & Arnold, 1988; Price & Liou, 1989; Alatalo, Gustafsson & Lundberg, 1990; Cooke et al., 1990). We will review these explanations, and describe methodological advances which can be used to distinguish among them. We illustrate these approaches using examples from our own recent work on wild populations of collared flycatchers *Ficedula albicollis* and red deer *Cervus elaphus*. It is our belief that it is as important to understand why evolution may be absent when it is apparently to be expected, as it is to understand why it is present.

Absence of evolution in the wild

A number of intensive studies of natural populations have documented both directional natural selection on a particular trait, and additive genetic variation underlying that trait, but have found no evidence that the mean phenotype is changing over time in the predicted direction (Table 1). A much larger number of studies have estimated both components, or one component under circumstances where the presence of the other can be assumed, without reporting whether the trait is changing over time or not. The classic example of a trait showing unexplained stasis is breeding time in birds (e.g., Price, Kirkpatrick & Arnold, 1988). In many temperate zone bird species, early breeding individuals have higher reproductive success than late breeding ones (e.g., Klomp, 1970; Findlay & Cooke, 1982; Daan, Dijkstra & Tinbergen, 1990; Merilä & Sheldon, 2000; but see van Noordwijk, van Balen & Scharloo, 1981), and experimental manipulations in several bird species suggest that most, if not all, of this selection is causally related to date of breeding, rather than reflecting selection on a correlated variable (e.g., Norris, 1993; Wiggins, Pärt & Gustafsson, 1994; Verhulst, van Balen & Tinbergen, 1995). Those studies which have possessed sufficiently extensive information about the relationships between individuals have also shown that a substantial proportion of the variance in breeding time is apparently due to additive genetic variance (e.g., van Noordwijk, van Balen & Scharloo, 1981; Findlay & Cooke, 1982; Merilä & Sheldon, 2000). Despite these two requisite parts of the 'breeders equation' (Eq. (1)) being non-zero, most populations of wild birds, some of which have been studied for more than 50 years (approximately 30 generations), show little sign of evolving towards earlier breeding times. In the discussion that follows we will often return to this example to illustrate our points: this is not meant to imply that we feel this is a more interesting or important case than others, rather we feel that it is convenient to keep one example under close scrutiny.

At this point it should be noted that numerous studies have reported changes in breeding phenology, in birds and other organisms, over the last few decades (e.g., Beebee, 1995; Forchhammer, Post & Stenseth, 1998; Brown, Li & Bhagabati, 1999; Hughes, 2000). In these cases, stronger correlations with climatic factors than with time (e.g., McCleery & Perrins, 1998) suggest that the observed changes are due to changing environments, rather than reflecting a steady microevolutionary response to selection for earlier breeding. It is possible that the population-level response observed in these cases could represent the population's microevolutionary tracking of

Table 1. Examples of vertebrate studies which have demonstrated strong directional selection on apparently heritable traits, but no selection response over generations

Species	Trait	h^2	s'	Expected response (haldanes)	Number of years studied	Observed change in direction predicted?	Reference[e]
Cervus elaphus	Antler mass	0.36	0.46	0.165	29	Opposite direction	1
	Birth mass	0.11 (M)[a]	0.40[b]	0.044		No change	
		0.25 (F)	0.22[b]	0.055		No change	
Ovis aries	Body mass	0.12 (M)[a]	0.11[b]	0.013	12	No change	2
		0.24 (F)	0.07[b]	0.017		No change	
Branta leucopsis	Tarsus length	0.53	0.030 (M)[a]	0.016	13	Opposite direction	3
			0.093 (F)[a]	0.049			
Anser caerulescens	Body size	0.50	Positive[c]	–	18	Opposite direction	4
	Clutch size	0.20	0.33	0.066	20	Opposite direction	5
Ficedula albicollis	Relative mass	0.30	0.23	0.069	17	Opposite direction	6
	Tarsus length	0.52	0.12	0.062	4	No change	7
	Tarsus length	0.35	0.18	0.063	17	No change	8
	Breeding time	0.19	0.31[d]	0.059	19	No change	9
Parus major	Breeding time	0.50	0.46	0.230	23	No change	10

[a]M = males; F = females.
[b]Mean of repeated selection events.
[c]Not given as selection gradient, but strong selection inferred from figures in Cooke, Rockwell and Lank (1995).
[d]Standardised selection gradient, controlling for clutch size which is genetically correlated with breeding time ($r_A = -0.43 \pm 0.11$).
[e]1. Kruuk et al. (2000, 2001) and Unpublished data, 2. Milner et al. (1999, 2000), 3. Larsson et al. (1998), Henk van der Jeugd, personal communication 4. Cooch et al. (1991), Cooke, Rockwell and Lank (1995), 5. Cooke et al. (1990), 6. Merilä, Kruuk and Sheldon (2001a,b), 7. Alatalo, Gustafsson and Lundberg (1990), 8. Kruuk, Merilä and Sheldon (2001), 9. Sheldon, Kruuk and Merilä (2001), 10. Boyce and Perrins (1987); Perrins and Jones (1974).
h^2 = heritability estimate of the trait, s' = standardized selection differential (i.e., selection differential in units of phenotypic standard deviations).

the changing environment, although the responses are generally too fine-grained to make this explanation particularly plausible. By this, we mean that the population shows such a close-tracking of the environment that the change cannot be entirely due to a microevolutionary response, given the degree of overlap of generations in most such cases. Thus it seems likely that these changes are actually driven by phenotypic plasticity. However, only one study has explicitly tested whether phenotypic plasticity can explain a population-level response to climatic variation. Przybylo, Sheldon and Merilä (2000a) showed, using longitudinal data from a population of the collared flycatcher, that the relationship (slope) between the annual value of the winter North Atlantic Oscillation index (NAO) and breeding date at the level of the population was the same as that estimated for individual birds breeding more than once, in years differing in NAO index. The same was true for clutch size (Przybylo, Sheldon & Merilä, 2000a). Hence, in this case, the entire response of the population could be attributable to phenotypic plasticity. More work is needed from other longitudinal studies to confirm whether this is generally the case, but the study nevertheless illustrates an important point: a population-level response consistent with microevolution need not reflect microevolutionary change, so that any claimed cases of change in observational studies must exclude the possibility that what is observed results solely from phenotypic plasticity (see James, 1983; Trussell & Smith, 2000 for illuminating examples of how this point also applies to studies of local adaptation). This caution might be thought to apply only to traits that are likely to display high degrees of behavioural plasticity, such as breeding time (or clutch size, or migrational direction). However, Wikelski and Thom (2000) showed that individual Galapagos marine iguanas apparently change their skeletal dimensions in response to environmental fluctuations (in this case fluctuations in the El Niño southern oscillation). Similarly, Losos et al. (2000) showed that hind-limb size in *Anolis sagrei* showed some phenotypic plasticity, implying that part of their earlier demonstration (Losos, Warheit & Schoener, 1997) of phenotypic change in response

to introduction to islands may not have involved a microevolutionary response. Hence, the class of traits which might be adjusted by individuals in response to climatic or environmental fluctuations may need to be expanded.

In conclusion, although there is good evidence that even rapid microevolutionary changes are of common place occurrence (e.g., Hendry & Kinnison, 1999; Bone & Farres, 2001; Kinnison & Hendry, 2001; Reznick & Ghalambor, 2001), numerous intensive long-term individual-based studies of natural populations have failed to detect evidence of ongoing evolution, and there are few cases where reported changes can be dissociated from a response to changing environmental conditions, either at the population or the individual level.

Approaches for detecting microevolutionary change

Before describing the strengths and weaknesses of long-term population studies of vertebrates for the study of contemporary microevolution, it is instructive to compare the pros and cons of different approaches to the study of microevolution in the wild. A simple classification of possible approaches to demonstrate the occurrence of microevolutionary changes with biometrical tools is presented in Table 2. According to this classification, microevolution can be detected either by direct observation from longitudinal (allochronic) data, or inferred from synchronic data (Table 2; terminology from Hendry & Kinnison, 1999). These two main approaches – henceforth referred as to direct and indirect approaches, respectively – can be further divided according to the type of methodology they employ.

Indirect approaches

Microevolution can be inferred from allochronic data in two different ways (Table 2). Either, one can search for consistent associations between trait means and selection pressures across different contemporary populations of the same species, or one can trace changes in character states in intraspecific phylogenies, and match these changes with changes in presumed selection pressures (Losos et al., 1998; Schluter, 2000). The latter approach has seldom been used in intraspecific studies (but see Edwards & Kot, 1995) and hence falls somewhat outside the scope of this paper. The first approach, on the other hand, is perhaps the most commonly used method to infer the occurrence of adaptive evolutionary changes (e.g., Endler, 1977; Berven & Gill, 1983; Linhart & Grant, 1996). However, although this approach can provide evidence that microevolution has occurred (if the genetic basis of the observed population differentiation can be demonstrated by common garden, transplant (e.g., Conover & Schultz, 1995) or hybridisation (e.g., Arnold, 1981; Caroll et al., 2001) experiments), its shortcoming is that the causality between postulated selection pressure and observed divergence can seldom be firmly established: although quite convincing cases can be made (e.g., Snakes: Arnold, 1981; *Drosophila*: Partridge & Coyne, 1997; Gilchrist & Partridge, 1999; Huey et al., 2000). Furthermore, although it may be possible to reject genetic drift in favour of directional natural selection as the explanation for the observed divergence (c.f. Spitze, 1993; Merilä, 1997; Lynch et al., 1999), the time-scale over which the divergence occurred will often remain unknown (but see Hendry & Kinnison, 1999). Consequently, although they may represent the only available means in many instances, indirect approaches seldom allow detailed inference about the pace and causes of microevolutionary changes in the wild.

Direct approaches

By direct observation we refer to a situation where a change over time in the mean value of a trait is actually measured within a population. Again, two basic approaches can be identified: experimental and observational (Table 2). The experimental approach aims to study the response of a population exposed to a novel selective environment. Examples of such approaches include the transplantation of guppies between streams differing in their predation regime (Reznick et al., 1997), the release of small groups of *Anolis* lizards to islands where no lizards were previously present (Losos, Warheit & Schoener, 1997) and the application of artificial selection in wild populations (Flux & Flux, 1982; Semlitsch & Wilbur, 1989). The direct approach based on observational data consists of documenting changes in the mean phenotype of a population of marked individuals with simultaneous estimation of forces of selection on and genetic variances of traits of interest. The best example of this approach is a study of Darwin's finches on the Galapagos (Grant & Grant, 1995).

Table 2. Classification of methods for detecting microevolutionary change in the wild

Approach	Method	Rationale	Example
Allochronic (direct)	Observational	Observe a change over time	Grant and Grant, 1995
	Experimental	Induce evolution over time	Reznick et al., 1997
Synchronic (indirect)	Observational	Selective regimes & trait means correspond across populations	Blondel et al., 1999
	Phylogenetic	Character state change implies evolution	Losos et al., 1998*

*The sympatric ecomorphs in this study are actually reproductively isolated, and hence, this may be an example of interspecific comparison. Nevertheless, it illustrates the principle how phylogenetic approach could be implemented in the current context.
In all cases, the implicit assumption is that changes or differences in trait means are genetically based (an assumption which can be tested e.g., Reznick et al., 1997), and that there is a causal relationship between trait mean and fitness.

Both experimental and observational approaches have their drawbacks. While an evolutionary response to a changed selective environment may be very informative, it is not always established that the observed change in the population mean reflects genetically based change (c.f. Losos, Warheit & Schoener, 1997; Losos et al., 2000). In addition, the forces of selection will rarely be quantifiable: two environments are likely to differ in so many ways that it will be very hard to identify the factor that caused the change. In another sense, the observation that experimental exposure to a novel environment causes microevolution is trivial: it does not actually tell us very much about the rates at which evolutionary divergence might have occurred in nature, or about the way in which unperturbed populations may or may not track their environments. Artificial selection in the wild, though rarely performed, is free from some of these criticisms. On the other hand, one could argue that it is in fact not far removed from laboratory-based artificial selection experiments. The main problems associated with the application of the observational approach are the assumptions that environmental influences affecting the traits of interest have remained constant over generations (e.g., Cooke et al., 1990; van Noordwijk, 1990), and that the relationship between fitness and the postulated selection pressure causing the observed change in the trait mean is causal (e.g., Dhondt, Eyckerman & Huble, 1979; Crespi, 1990). In addition, because the forces of selection in natural populations may be weaker than those caused by an experimental perturbation, a purely observational study may need both vast sample sizes and very long time series in order to have any power to detect an equivalent evolutionary response. Hence, both of the direct (allochronic) approaches are likely to be forced to investigate shorter time series than the indirect (synchronic) approaches, and may therefore be biased towards exploring the outcome of strong natural selection.

Before shifting the focus on long-term studies of individually marked animals, we wish to emphasise the fact that we do not consider the evidence for microevolution obtained from other approaches (c.f. Table 2) to be less valuable than that which can be obtained from observational studies of individually marked populations. Instead, our reason to focus on direct observational studies is in the fact that they allow, at least in principle, a rather uncomplicated application of the quantitative genetic models (Eqs (1) and (2)) of microevolution to actual data. This is seldom possible with the other approaches unless assumptions are made about ancestral character stages, genetic parameters and/or forces of selection acting on traits of interests.

Inferring microevolutionary changes from long-term studies

In a typical long-term field study, information is collected about reproduction, survival and morphology of marked individuals over a long period of time, and consequently, the forces of selection acting on different traits are, at least in principle, estimable (Lande & Arnold, 1983; Arnold & Wade, 1984a,b). In many long-term studies information about relatedness among individuals, critical for the estimation of quantitative genetic parameters of evolutionary interest, is also collected as a by-product of the marking scheme. Consequently, given that the data required to estim-

ate heritability (or **G**) and the magnitude of selection acting on a given trait are often readily available, and that these are the very elements needed to predict the evolutionary response to selection as given by Equations (1) and (2), studies of this kind should be valuable for exploring the presence or absence of microevolution. However, many studies that have collected data of this sort have not used it extensively to estimate selection and quantitative genetic parameters, let alone to try and combine the two. The reasons for this are not clear, but part of the explanation may lie in the fact that many long-term studies were set up by biologists interested in testing ecological theories of reproductive scheduling (e.g., testing Lack's (1968) model of clutch size). Another reason for the paucity of attempts to explore microevolutionary changes in long-term studies may be that counter-intuitive or otherwise problematic results from early studies (e.g., Dhondt, Eyckerman & Huble, 1979; Cooke et al., 1990; Alatalo, Gustafsson & Lundberg, 1990) may have discouraged further exploration of this topic.

Table 1 lists those longitudinal studies that we are aware of that have tested whether the population mean phenotype is changing in a manner consistent with microevolution, over more than one generation. The notable point about these studies is that none of them report a change in the expected direction; five even report a change in the opposite direction to that expected on the basis of selection gradients and quantitative genetic parameters (Table 1). These studies have been conducted on populations that range from closed ones, where most or all of the population can be studied (Milner et al., 1999, 2000; Kruuk et al., 2001), to those which form part of a large continent-wide metapopulation (e.g., Cooke et al., 1990). The traits studied include both morphological characters and life history characters, with one case of a sexually selected character (Kruuk et al., 2001). The one thing that all of these cases have in common is that the selection documented on the phenotypic values of the traits is strong, exceeding the median strength of selection in natural populations estimated by Kingsolver et al. (2001) in all but two cases. In most cases, the sample sizes are large (in some cases, huge: e.g., $> 20\,000$ in Merilä, Kruuk & Sheldon, 2001a, Kruuk, Merilä & Sheldon, 2001), and these studies thus represent some of the strongest cases of natural selection observed for their sample sizes (Kingsolver et al., 2001). Several hypotheses have been proposed to explain the absence of a microevolutionary response in such cases.

Explanations for lack of selection response

1. *Biased estimates of heritability*

An obvious explanation for the lack of selection response is that one of the elements of Equations (1) or (2) has been wrongly determined. Estimates of heritability (or additive genetic variance), which in most vertebrate studies are derived by comparing the resemblance between parents and offspring with parent-offspring regressions, can be inflated by environmental covariance between parents and offspring or by the presence of maternal effects (Bernardo, 1996; Falconer & Mackay, 1996; Mousseau & Fox, 1998). Fifteen cross-fostering experiments conducted with birds provide little evidence that heritability estimates are seriously biased by environmental covariances between offspring and parents (reviewed in Merilä & Sheldon, 2001). However, since cross-fostering experiments are performed in particular years, and typically over relatively short spatial distances, they do not necessarily control for all possible environmental covariance among parents and offspring. Recent studies based on an 'animal model' approach (see below) have shown that maternal effects may account for a significant component of the phenotypic variance in a trait (e.g., Kruuk et al., 2000; Milner et al., 2000; Coltman et al., 2001), such that estimates of heritability from mother-offspring regressions will be biased upwards if they are ignored (see Milner et al., 2000 for an explicit comparison of heritability values estimated with and without considering maternal effects). Similarly, substantial common-environment effects will inflate similarities between siblings, again with the results that estimates of heritability using full or half-sib analyses would be inflated if these were not taken into account (e.g., Kruuk, Merilä & Sheldon, 2001; Merilä, Kruuk & Sheldon, 2001a). Hence, although these biases are possible, their likely effect is to slow down the selection response relative to that expected on the basis of Equations (1) and (2), rather than to lead to a situation where the response is lacking or occurs to direction opposite to that predicted. In the same vein, heritability estimates may be biased by incorrect assignment of offspring to their parents, which happens easily due to uncertainty of paternity in studies of free-living populations. However, this source of bias is an unlikely explanation for the observed mismatch between expected and observed selection responses as the extra-pair paternity should deflate, rather than inflate, the estimates of heritability.

Quantitative genetic models suggest that indirect (genetic) maternal effects can have interesting consequences for the prediction of evolutionary responses at the level of the phenotype. Depending on the covariance between additive genetic and maternal genetic effects, the response to selection may be either accelerated or retarded (e.g., Kirkpatrick & Lande, 1989; Roff, 1997; Shaw & Byers, 1998; Wade, 1998). In fact, under certain circumstances, maternal effect influences could indeed explain selection responses not predictable from standard quantitative genetic equations (e.g., Roff, 1997, p.251). Hence, the possible role of maternal or common environmental effects in explaining the mismatch between expected and observed selection responses needs to be investigated in more detail, with more attention paid to the covariance between additive genetic and maternal genetic effects.

2. *Selection fluctuates in time or space*

The other key component in the quantitative genetic models of evolutionary change (Eqs. (1) and (2)) is the strength of directional selection acting on the trait. By using a single value to represent the strength of selection, an implicit assumption is that the strength of selection remains constant over time. If, on the other hand, selection fluctuates widely from year to year, this may act to slow down any expected evolution (e.g., Price & Liou, 1989). In addition, all of the organisms in Table 1 have overlapping generations; selection, however, tends to be estimated on a year-by-year basis. Predictions about the expected response to selection should be couched in terms of the rate per generation, which is not always known with certainty in wild populations. For example, Hendry and Kinnison (1999) point out that Endler (1980) and Reznick et al. (1997) used estimates of generation time in guppies that differed by a factor of five, despite working with the same population. This may be of particular consequence if the age-dependent reproductive schedule differs between the sexes, as may the case for strongly polygynous mating systems. Relatively few studies have investigated how selection changes from year to year (or generation to generation) within the same population. In some cases, selection intensity and direction may fluctuate widely from year to year (e.g., Price et al., 1984; Grant & Grant, 1993). In other cases, selection varies little between years (e.g., Boyce & Perrins, 1987; Przybylo, Sheldon & Merilä, 2000b; Kruuk, Merilä & Sheldon, 2001; Sheldon, Kruuk & Merilä, 2001). Here, it also worth pointing out that incorrect estimation of selection intensity could also occur if selection varies through the life history of the organism and is only measured for a part of life history (Schluter, Price & Rowe, 1991). Equally, incorrect estimation of selection intensity could occur if a trait contributes in multiple ways to fitness and only one fitness component is measured. As to the latter point, the strength of direct (allochronic) approaches is that estimates of total life-time reproductive success are possible to obtain (e.g., Kruuk et al., 2000; Merilä & Sheldon, 2000), and this source of error can be minimised.

Another source of fluctuating selection pressures may be spatial heterogeneity. For example, different populations of a species may be selected in different directions with respect to the same environmental factor, but if they exchange many migrants, the evolutionary response in both directions may be constrained (e.g., Hendry, Day & Taylor, 2001). This process has been proposed as a constraint on the evolution of clutch size in great tits *Parus major* inhabiting patchy landscapes in Europe (Dhondt et al., 1990), the evolution of antipredator behaviour in salamanders (Storfer & Sih, 1998), and local adaptation in spiders (Riechert, 1993). Similarly, dispersal between evergreen and deciduous woodlands, which have different optimal breeding dates, has been suggested to constrain the evolution of breeding time in blue tits *Parus caeruleus* (Dias & Blondel, 1996). Gene flow from other populations of the same species cannot, of course, offer an explanation for the absence of an evolutionary response in insular populations (e.g., Milner et al., 1999; Kruuk et al., 2001), nor can it explain stasis when individuals from different populations are exposed to the same environmental conditions, such as viability selection acting during avian winter migrations (Kruuk, Merilä & Sheldon, 2001; Merilä, Kruuk & Sheldon, 2001). However, another source of gene flow may be due to hybridization with closely related species (Barton & Gale, 1993), a process which occurs in both the Darwin's finches on the Galapagos, and in the collared flycatcher *Ficedula albicollis* population studied on Gotland, Sweden (see Table 1). Little is known about the rates of introgression at quantitative trait loci in natural populations, but theoretical models suggest that both in this case, and in the case of gene flow between populations of the same species, rates of gene flow have to be rather high to counteract local selection (e.g., Lande, 1980; Garciá-Ramos & Kirkpatrick, 1997; Hendry, Day & Taylor, 2001). However, the effect of gene flow on

adaptation will depend not only upon the rate of population mixing, but also on the amount of additive genetic variance in the trait and the degree of difference in trait means between mixing populations (Hendry, Day & Taylor, 2001). Hendry, Day and Taylor (2001) provide a useful framework to evaluate the role of gene flow in constraining adaptation with worked examples from Lake Erie water snakes (*Nerodia sipedon*) and sticklebacks (*Gasterosteus aculeatus*).

3. *Selection on environmental deviations*

Fisher (1958) pointed out that, even if a trait is heritable, it may not respond to selection if natural selection acts predominantly on the non-heritable component of the phenotype. This point has been repeatedly invoked as an explanation for stasis ever since (e.g., Alatalo, Gustafsson & Lundberg, 1990; van Tienderen & de Jong, 1994), although the phenomenon has usually been framed in terms of 'misidentified target of selection' (e.g. Price, Kirkpatrick & Arnold, 1988; Price & Liou, 1989). One of the most influential treatments was that of Price, Kirkpatrick and Arnold (1988), who used this reasoning to explain the lack of evolutionary response to selection on breeding time in birds. In short, assume that we have two phenotypically correlated traits (traits 1 and 2 in Eq. (3)), one ('condition'; trait 2) which is a true target of directional selection ($\beta > 0$) and another (laying date; trait 1) which is not ($\beta = 0$). Assume further that laying date is heritable ($G_{11} > 0$) but condition is not ($G_{22} = 0$), and hence that there is no genetic covariance among these traits (i.e., $G_{12} = G_{21} = 0$). In this situation Equation (3) predicts that we should not expect selection response in either of these traits, in spite of the fact that the univariate equation (Eq. (1)) would indicate so. The reason for this can be illustrated with a simple path diagram depicting the essential features of the Price–Kirkpatrick–Arnold model (Figure 2). Their argument was that it was likely that non-heritable condition, or nutritional state, would cause individuals both to breed earlier and to have higher reproductive success, hence generating a correlation between breeding time and fitness (Figure 2). If condition is not measured (in which case the example reduces to univariate situation), then it would appear that there is causal correlation (i.e., selection on laying date) between phenotype (breeding time) and fitness (r_{pw}), even if there were none (Figure 2). The argument can be extended to any other phenotypic trait that is under selection (e.g., clutch size: Price & Liou, 1989; antler size: Kruuk et al., 2001; see also Rausher, 1992), and this is a general and well-known problem with studies of natural selection: selection on unidentified (and/or unmeasured) traits can seriously compromise our attempts to predict selection responses both in laboratory and wild.

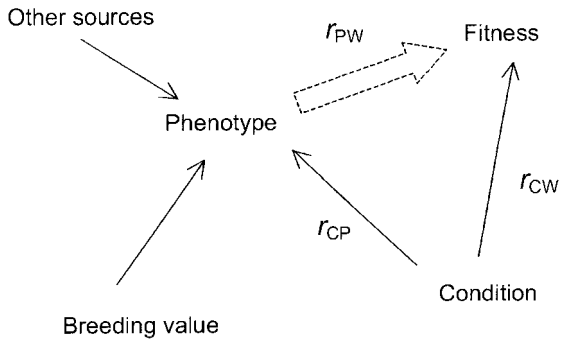

Figure 2. Price–Arnold–Kirkpatrick (1988) model to explain the absence of evolution of quantitative traits due to selection acting on environmental deviations of traits. The fat arrow (r_{PW}) is the correlation between the phenotype and fitness resulting from the joint effect of condition (or nutritional state) on the phenotype and fitness (r_{CP} and r_{CW} respectively).

Although we acknowledge that the Price–Kirkpatrick–Arnold model provides a useful theoretical and conceptual framework to explain lack selection response when such is expected, its practical utility is limited. There are two main reasons for this. Firstly, by definition, it is hardly ever possible to be sure that selection is acting on those traits we measure, and not on some phenotypically correlated unmeasured trait(s). Secondly, in the example described above, the genetic covariance between the true target of selection and the correlated trait has to be zero – otherwise at least some correlated selection response would be expected (cf. Eq. (3)). The importance of the latter point should be obvious: how can we ever estimate the genetic correlation between two traits of which one is measured and the other is not?

An alternative way to view the problem is to rephrase it terms of breeding values and environmental deviations (e.g., Lynch & Walsh, 1998). Using the Price–Kirkpatrick–Arnold example of laying date to illustrate our point, we note that an individual's phenotypic breeding date (P) is assumed to be a sum of its breeding value for laying date (A) and some random environmental deviation (E) from this value (i.e., P = A + E). Hence, some individuals will be early breeders because of their large negative environmental deviations in respect to laying date. Although these environmental deviations are random with

respect to breeding values, the underlying cause for negative (or positive) environmental deviations need not be so: negative environmental deviations representing early breeding could be attributable to some unmeasured factor such as condition. The core of this reasoning is that if we have a means of estimating individual breeding values and environmental deviations, we can re-phrase the problem of 'selection on unmeasured traits' in terms of selection acting on genetic and environmental components of variance in a given trait. However, as pointed out by Endler (1986, see also Milkman, 1982; Lynch, 1984), very little is known about the degree to which selection acts on breeding values or environmental deviations in natural populations.

There are two ways in which the idea that selection acts primarily on environmental deviations could be tested, where we define the environmental deviation as the residual from a regression of individual phenotypes on individual breeding values. First, the size of the additive genetic (V_A) and environmental components (V_E) of variance before and after selection can be compared. If selection is acting on V_E only, then we would expect a significant reduction in V_E with each round of selection, but no reduction in V_A (Kruuk, Merilä & Sheldon, 2001; Merilä, Kruuk & Sheldon, 2001a). Alternatively, one can estimate the actual breeding values (e.g., Lynch & Walsh, 1998) of the individuals under selection, and quantify the forces of selection acting on breeding values, as suggested by Rausher (1992). In the presence of environment-driven correlations between the trait in question and fitness, the selection differential on phenotype will be greater than on breeding values (Rausher, 1992). Note that this is equivalent to an upwardly-biased estimate of the effective selection differential S in the Equation (1), which will therefore generate misleading expectations of evolution in the same way as inflated heritability estimates. However, if there is a significant selection differential for breeding values, selection on environmental deviations is not, on its own, a sufficient explanation for the lack of any selection response in the trait.

4. *Selection on correlated traits*

Another potential reason for the lack of expected selection response in a given trait is that genetic correlations between it and other traits may constrain the pace and direction of evolution (e.g., Lande & Arnold, 1983; Price & Langen, 1992; van Tienderen & de Jong, 1994; Barton & Partridge, 2000). For example if trait 1 in Equation (3) is selected in the same direction as trait 2, but the two traits are negatively genetically correlated, the expected selection response in trait 1 is not only a function of strength of selection acting on trait 1, but depends also on (i) the strength of the genetic correlation between trait 1 and 2, and (ii) the strength of selection acting on trait 2 (Lande & Arnold, 1983; Eq. (3)). To give a more concrete example, laying date and clutch size often show a strong negative phenotypic correlation in birds, such that later laying birds lay smaller clutches (e.g., Wiggins, Pärt & Gustafsson, 1994). If clutch size is subject to stabilising selection, and the negative phenotypic correlation indicates an underlying negative genetic correlation, then selection for earlier breeding would be counteracted by the stabilising selection on clutch size, since genes for breeding early would tend to be associated with laying larger clutches, further away from the optimum for clutch size. Recent work by the authors on the collared flycatcher shows that the phenotypic correlation between laying date and clutch size ($r_P = -0.38$, $N = 4852$, $p < 0.0001$) is partly due to a negative genetic correlation between these traits ($r_A = -0.44 \pm 0.10$, $p < 0.001$), suggesting that selection on clutch size would have the potential to constrain any evolution of laying date (Sheldon, Kruuk & Merilä, 2001). However, multiple regression analysis suggests that there is very little selection on clutch size, once its dependence on laying date is accounted for, implying that counter-selection on clutch size is in fact unlikely to constrain any evolutionary response in laying date (Sheldon, Kruuk & Merilä, 2001; see also Boyce & Perrins, 1987).

One of the difficulties with the suggestion that genetic correlations may constrain evolutionary responses is that it is impossible to be sure that one has measured all traits that might potentially be genetically correlated with the focal trait (Lande & Arnold, 1983; Crespi, 1990; see above). This is particularly the case for life-history traits, for which genetic correlations with other life history traits are expected, even if their sign is not predictable (Houle, 1991). Grant and Grant (1995) demonstrated the utility of knowing the genetic variance-covariance structure (**G**) for multiple traits in their study of evolution of bill size in Galapagos finches, where the predicted response of some traits was displaced in the opposite direction from that which would have been predicted on the basis of univariate phenotypic selection alone. However, the utility of predicting evolutionary responses

from known **G** may be limited over relative short time frames, as **G** is also expected to evolve in response to selection (e.g., Roff, 2000). Nevertheless, inferences based on **G** even when the time-frame is longer appear to have some utility (e.g., Merilä, Björklund & Gustafsson, 1994; Schluter, 1996, 2000; Merilä & Björklund, 1999).

A further form of genetic correlation can also constrain evolutionary responses: this occurs when there is a genetic correlation between the sexes for a trait, and selection acts in opposite directions in the two sexes. This type of sexually antagonistic selection has been recently demonstrated in studies of two passerine birds (Price, 1996; Merilä, Sheldon & Ellegren, 1998), and since genetic correlations between the sexes are typically very high (Roff, 1997, p. 247), there is reason to expect this constraint to be wide-spread. Again, sexually antagonistic selection acting on breeding values rather than environmental deviations (or environmental deviations in one and on both the genetic and environmental components in the other sex) has not to date been investigated in wild populations (but see Kruuk, Merilä & Sheldon, 2001). Another issue worth pointing out in this context is that although it may be tempting to measure selection in one sex only (e.g., for a secondary sexual character), one should not ignore the fact that genes for the trait will also pass through the opposite sex (unless Y/W linked) and could potentially have effects in that sex too, even if the focal trait is apparently not expressed.

5. *Genetic response masked by a changing environment*

An assumption implicit in any study using longitudinal data to investigate microevolutionary change over time is that the environment where the performance of the individuals is tested has remained constant over time. If the environment has changed at the same time as the presumed evolutionary changes have taken place, then it is possible that the genetic response to selection may be masked by opposite changes in the environment, or even worse, that the change in the mean phenotype of the populations is solely due to changes in environmental component of variation in the given trait (Cooke et al., 1990). It is instructive to note that this situation is exactly parallel to that frequently documented in studies of geographic differentiation and termed 'counter-gradient variation' (Conover & Schultz, 1995). Counter-gradient variation is formally defined as a negative covariance between environmental and genetic influences on a trait (Conover & Schultz, 1995), which, when it occurs, tends to hide spatial genetic divergence among populations due to non-random distribution of genotypes across environments. This occurs, for instance, when fast growing genotypes inhabit cold environments which tend to slow down growth rates and vice versa (see Conover & Schultz, 1995 for a lucid treatment of the phenomenon and arguments as to why this form of genotype-environments associations is expected to be common).

Returning to the temporal dimension, Cooke et al. (1990) argued that this type of process may have been responsible for the absence of an evolutionary response to selection on clutch size in lesser snow geese *Anser caerulescens.* They reasoned that while selection favoured individuals that laid larger clutches ($S = 0.33$), and clutch size was heritable ($h^2 = 0.20$), the increased recruitment of young geese from the successful birds would tend to increase population density. If the population were restricted to a limited area, then negative density-dependent processes would tend to decrease the environmental component of clutch size, despite the fact that the mean breeding value was increasing. Cooke et al.'s suggestion is appealing because it may be generally applicable to populations inhabiting a limited range: selection on a trait implies that some individuals with particular values of the trait have higher reproductive success, and produce more offspring, which tend to resemble their parents. These offspring inhabit the same range that their parents did, and are subject to the same form of selection themselves. As the population density rises, the share of the limited environment that each individual has access to decreases, with the result that any tendency for the phenotype to increase because of an increase in the population's mean breeding value is counteracted by a decline in the environmental component of the phenotype. There is no reason to expect that changes in the breeding value and environmental deviation should cancel each other out exactly. In the case of the population of barnacle geese *Branta leucopsis* studied by Larsson et al. (1998), a decline in the population mean, despite selection in the opposite direction, was attributed to an increased competition for food due to rapidly expanding population.

6. *Lack of statistical power*

Finally, as with all studies of natural populations, conclusions may be hampered by a lack of statistical

power, with significant biological effects being swamped by the unavoidable sampling error associated with relatively small sample sizes. For example, Kingsolver et al. (2001) suggest that the majority of studies of selection in the wild have lacked the power to detect selection of even an 'average' magnitude (the median sample size in their reviewed studies being $N = 134$). Any apparent stasis or lack of evolution may be merely due to consideration of too short a window in the population's evolution, or too great uncertainty in the population estimates to detect change. Amongst the studies that have tested explicitly for a response to apparent selection (e.g., Larsson et al., 1998; Milner et al., 1999; Kruuk et al., 2001), none has considered the power with which they might reject a null hypothesis of no change, nor whether there would be sufficient power in the data to detect change of the predicted magnitude. Even if there is indeed an absence of any microevolutionary change because of one (or more) of the above reasons, sufficient power to accept any of the explanations will also be required. Multiple tests (e.g., of linear and quadratic selection differentials and gradients on several traits) may also generate false expectations as to the significance of observed selection on a trait, and hence of the expected response (Kingsolver et al., 2001).

However, in considering a non-significant response R, when h^2 and S (Eq. (1)) have both been estimated as being statistically significant, it should be noted that trait values are typically easier to estimate than either heritabilities or selection differentials: phenotypic measurements are usually available for a larger sample of individuals in a population than are data on pedigree or fitness. It, therefore, seems unlikely that a study should return significant estimates of both heritability and selection differentials and yet fail to observe a phenotypic response simply because of the low statistical power. Furthermore, by their nature, the long-term studies are less likely to fall foul of the sample size problem.

Distinguishing the hypotheses

Of the six explanations that we have outlined above for why a population may show stasis, rather than evolution, despite the apparent presence of directional selection on a heritable trait, there has been little positive evidence to support any one of them in a specific case in a wild population, until very recently. One reason for this is that a number of the hypotheses (particularly numbers 3 and 5) are explanations based on quantities (individuals' environmental deviations and breeding values) which have not, traditionally, been estimated by field biologists. Some individual cases have been offered as support for one or other of the explanations. For example, Alatalo, Gustafsson and Lundberg (1990) analysed selection on measures of fledgling collared flycatchers, and concluded that apparent selection on body size (tarsus length) was actually mediated through selection on relative body mass ('condition'). Since they assumed that relative body mass reflected nutritional status only, they equated it with the environmental deviation of an individual's phenotype. However, recent work with a substantially larger sample of the same population has come to two conclusions at odds with those of Alatalo, Gustafsson and Lundberg (1990). First, there is evidence that tarsus length is the target of natural selection independently of relative body mass (Kruuk, Merilä & Sheldon, 2001, Table 1), and secondly there is good evidence for a substantial additive genetic component to relative body mass (Merilä, Kruuk & Sheldon, 2001, Table 1). Furthermore, one should of course note that the above explanations need not be exclusive: evidence in support of one may not be the full story, and indeed it is plausible that more than one of the processes/factors may apply simultaneously.

Animal model analyses

Developments in animal breeding science have provided a powerful analytical framework for the quantitative genetic analysis of fragmentary and unbalanced data, based on restricted maximum likelihood (REML) estimation of variance components and a mixed model analysis (reviewed in Lynch & Walsh, 1998, see also e.g., Groeneveld & Kovac, 1990; Groeneveld 1995; Meyer 1991, 1997). In particular, the 'animal model' approach expresses the phenotype of an individual in terms of its additive genetic merit, other random effects (such as the maternal or common environment effects discussed above) and fixed effects (such as e.g., age or sex differences):

$$y_i = \mu + a_i + u_{i1} + u_{i2} + \cdots b_{i1} + b_{i2} + \cdots + e_i, \quad (4)$$

where y is the phenotype of individual i, μ is the population mean, a_i is the individual's additive genetic value, u_{ij} are other random effects, b_{ij} are fixed effects and e_i is a random residual value (Knott et al., 1995;

Table 3. Potential impact of the application of 'animal model' parameter estimation to explanations of evolutionary stasis in natural populations

Explanation	Impact of application of 'animal models'
1. Inaccurate estimation of parameters	More accurate and less biased estimation of additive genetic and maternal effect variances
2. Fluctuating selection	Fluctuating selection on BVs testable; comparison of BVs of residents versus immigrants
3. Selection on environmental deviation	Direct estimation of both breeding value and environmental deviation
4. Selection on correlated traits	Improved ease of estimation of genetic correlations
5. Environment masks genetic response	Direct estimation of both breeding value and environmental deviation
6. Lack of statistical power	More efficient and extensive utilisation of data

BV = Breeding value.

Lynch & Walsh, 1998). The random effects typically specify factors such as maternal identity or common environment, and thus can account for additional correlations between relatives above those due to additive genetic effects. Fixed effects typically specify factors such as year of measurement or age of the individual. Animal model analyses utilise all the available information in multigenerational pedigrees, exploiting links between an individual and all other individuals to which it is related. They are therefore significantly more powerful than less complex approaches such as parent-offspring regression. They also accommodate unbalanced data sets and hence the missing values typical of data collected on natural populations: for example, phenotypic measurements on a grandmother and grandchild can still be used even without measurement of the mother's phenotype. The REML analyses provide estimates of components of variance in a base population that are unbiased by any effects of finite population size, selection or inbreeding in subsequent generations (Sorenson & Kennedy, 1984; van der Werf & de Boer, 1990). Because information in any pedigree rarely dates back to a true base population, an assumption concerning the base population is usually made, namely that the first generation of animals with data form the base population; the subsequent analysis will then estimate the components of variance in this first generation. In addition, and most significantly for the issues outlined above, animal models can provide estimates of individual breeding values, the expected effect of the genes that an individual passes on to its offspring (Falconer & Mackay, 1996; Lynch & Walsh, 1998), which can then be used to test for genetic trends in response to selection (e.g., Meyer & Hill, 1991; Southwood & Kennedy, 1991; Ferraz & Johnson, 1993; Martinez, Bünger & Hill, 2000, see also Hill & Cabellero, 1992; Ollivier, 1999 for discussion of the sensitivity of parameter estimates using animal models). Finally, they readily allow multivariate analyses and straightforward estimation of genetic correlations and associated standard errors (e.g., Groeneveld, 1995).

Despite these advantages and their widespread use in animal breeding studies, REML-based animal models have not yet been widely exploited in evolutionary studies of natural populations. To our knowledge, they have been applied in the analysis of data from long-term studies of only four wild animal populations to date, all very recent: bighorn sheep (*Ovis canadensis*) on Ram Mountain, Alberta (Réale, Festa-Bianchet & Jorgensen, 1999); Soay sheep (*Ovis aries*) on St Kilda, Scotland (Milner et al., 2000; Coltman et al., 2001); red deer (*Cervus elaphus*) on the Isle of Rum, Scotland (Kruuk et al., 2000, 2001) and collared flycatchers (*Ficedula albicollis*) on the island of Gotland, Sweden (Kruuk, Merilä & Sheldon, 2001; Sheldon, Kruuk & Merilä, 2001; Merilä, Kruuk & Sheldon, 2001a,b). These techniques offer several opportunities for testing the explanations for stasis in wild populations outlined above, some of which have not been available previously (Table 3). In the rest of this paper we will illustrate recent applications of these techniques to data from our work on collared flycatchers and red deer. In these cases, pedigree information is available up to a maximum of 13 and seven generations respectively, but animal models can be applied to data from considerably fewer generations (e.g., Brotherstone, McManus & Hill, 1990; Knott et al., 1995).

1. *Body size and condition in fledgling collared flycatchers*

Over 20 years of intensive monitoring of a population of collared flycatchers in the south of the Baltic island of Gotland, a total of more than 30 000 individuals have been ringed as fledglings. Approximately 12% of these returned to the study site as breeding adults, so that measurements on fledglings can be combined with pedigree data to estimate relevant quantitative genetic parameters. Estimates of the components of variance in tarsus length (Kruuk, Merilä & Sheldon, 2001) and condition (Merilä, Kruuk & Sheldon, 2001a) revealed a substantial component of additive genetic variance for both traits, corresponding to significant estimates of heritability (Table 1). However, despite strong viability selection on the two traits, such that both tarsus length and condition were significant predictors of whether an individual survived to breeding age (Table 1), there was no evidence of any increase in the mean phenotypic value of tarsus length, and evidence of a decline in mean condition index over the study period (Table 1). Using the results from animal model analyses of the data, the first three of the above hypotheses can be eliminated as possible explanations for this lack of microevolution. First, inclusion of a random effect of nest-box identity prevented heritabilities being over-estimated due to common environment or maternal effects – and also revealed that 30% of the phenotypic variation in tarsus length and 49% in condition was due to common environment and/or maternal effects, above that due to additive genetic variance (Kruuk, Sheldon & Merilä, 2001; Merilä, Kruuk & Sheldon, 2001a). Second, there was no suggestion that the direction of selection on either trait fluctuated to a large extent between years (Figures 3(a),(b)). Finally, two lines of evidence indicated that selection was acting on the heritable component of variation. Selection differentials on individual breeding values were significantly positive for both traits in most of the study years (Figures 3(c),(d)) and across all years (tarsus length: $S = 0.13 \pm 0.02$ S.E., condition: $S = 0.14 \pm 0.02$ S.E.). Furthermore, when the variance component estimation procedure was restricted to the surviving individuals that returned to the study site as breeding adults, estimates of both the additive genetic variance and the environmental variance were reduced, confirming that selection was acting on both the heritable and the non-heritable component of phenotype (Figures 3(e),(f)). Hence, the lack of support for any of the proposed explanations for evolutionary stasis calls for further analyses, but at least the first three of the main contenders (explanations 1–3) can be eliminated with reasonable confidence.

2. *Antler size in red deer*

A recent study of the red deer population on the Isle of Rum, Scotland has provided evidence in support of the suggestion that apparent selection on a phenotypic trait may be concentrated on its non-heritable component, or environmental deviation (Price, Kirkpatrick & Arnold, 1988; see explanation 3 above). Red deer are polygynous, highly sexually-dimorphic ungulates, in which males grow antlers each year. Antlers are involved in intrasexual competition for mates, and determine fighting success and dominance rank. REML analysis of dry antler mass returned an estimate of heritability of 0.36 ± 0.06 S.E. (Kruuk et al., 2001). Taking measures of an individual's lifetime breeding success, defined as the number of calves that he fathered in his lifespan, as an estimate of an individual's fitness, there was evidence of strong selection on antler size – with significant positive correlations between an individual's average antler size across his lifespan (after correcting for age-related variation) and his fitness (Figure 4(a), $p < 0.001$). Combined, these results predict that antler size should have been increasing in the population, at a rate of 0.165 haldanes (or standard deviations per generation; see Table 1). However over the 29 year study period, or a period of approximately four generations, antler size had actually decreased in the Rum population (Clutton-Brock & Albon, 1989; Kruuk et al., 2001). Analysis of individual breeding values offers an apparent explanation: in contrast to the flycatcher studies described above, there was no evidence of a significant association between breeding values and fitness (Figure 4(b), $p = 0.09$), but strong selection on the environmental deviations (Figure 4(c), $p < 0.001$). Antler size is heavily dependent on nutritional state, so the reduction in mean phenotype presumably reflected decreased resource availability associated with rising population density in the study area, and hence a decreasing environmental component. The example thus illustrates the potential for analysis of breeding values to quantify the extent to which changes at the phenotypic level represent changes in the underlying genetic composition of a population.

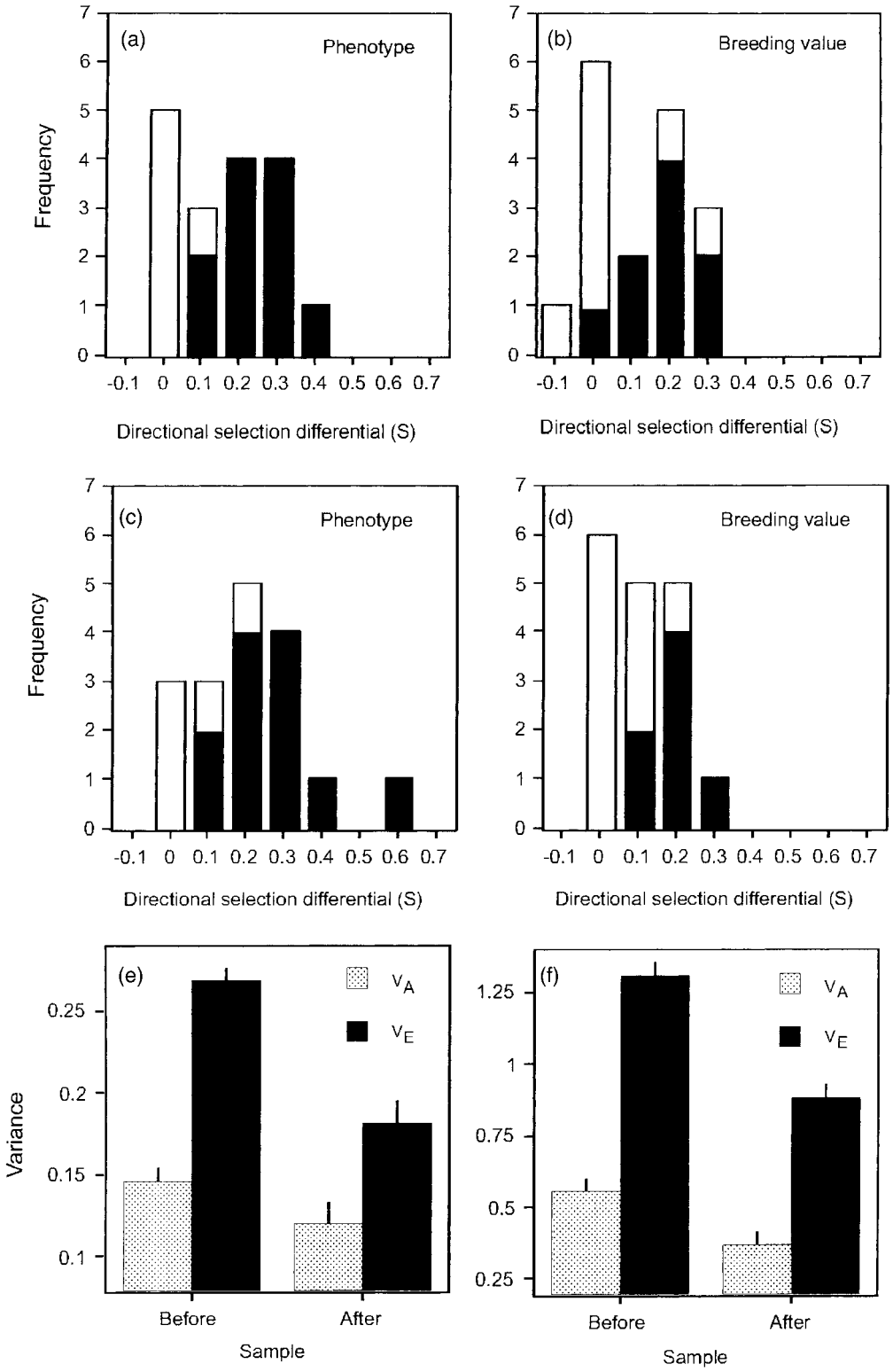

Figure 3. Selection on tarsus length and condition in fledgling collared flycatchers during 17 years. (a) Directional selection differentials on phenotypic values of tarsus length and (c) condition, (b) directional selection differentials on breeding values of tarsus length and (d) condition. 1981 = Year 1. Components of variance before and after selection for (e) tarsus length and (f) condition: V_A = additive genetic variance; V_E = total phenotypic variance minus V_A. In all figures, error bars represent one S.E. Data from Kruuk, Merilä and Sheldon (2001) and Merilä, Kruuk and Sheldon (2001a).

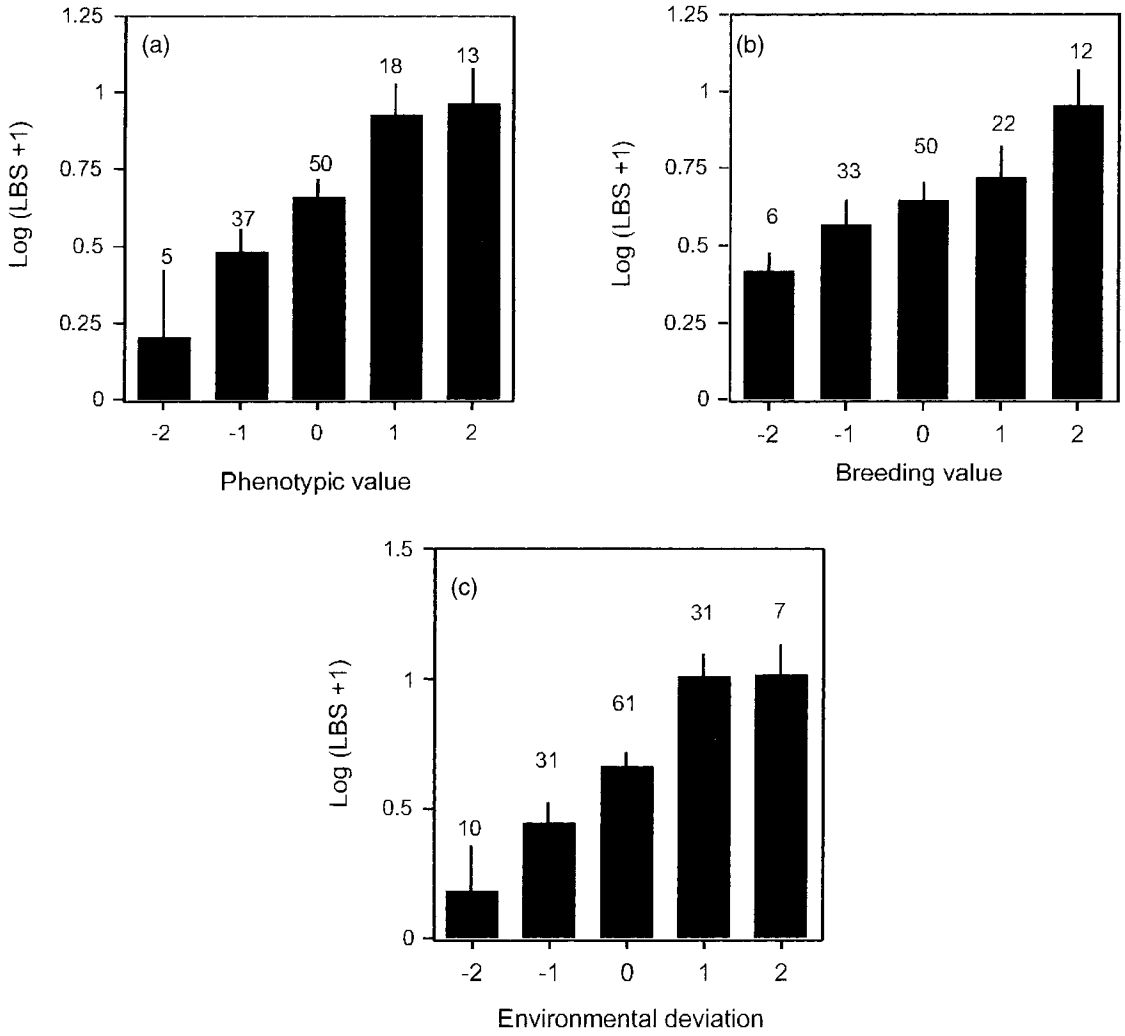

Figure 4. Selection on antler size in red deer. Graphed is a logarithm of lifetime breeding success (LBS +1) against (a) phenotypic value, (b) breeding value and (c) environmental deviation of antler weight. Values on x-axis are one standard deviation unit bins. Redrawn from data in Kruuk et al. (2001).

3. *Selection on laying date in the collared flycatcher*

One trait which has been much studied in the context of microevolutionary stasis is breeding time in birds. As discussed previously, there are numerous studies showing that earlier breeding tends to be favoured, because these breeding attempts are likely to recruit more offspring (Daan, Djikstra & Tinbergen, 1990; Merilä & Sheldon, 2000). A similar pattern has been demonstrated in several other taxa (e.g., mammals: Clutton-Brock et al., 1987, reptiles: Sinervo & Doughty, 1996; fish: Schultz, 1993). In birds breeding in temperate areas, the effect of date of birth on offspring fitness is often believed to be due to a corresponding decline in the availability of food for developing offspring (Daan, Djikstra & Tinbergen, 1990; Verhulst, van Balen & Tinbergen, 1995). Since selection on breeding time in birds is often strong, and there is evidence for reasonable amounts of additive genetic variance in breeding time (see above), this is an ideal trait in which to ask whether there is any evidence of an evolutionary response to selection. We used data from a long-term study of the collared flycatcher, to attempt to answer this question (see Merilä & Sheldon, 2000 for more details of this population, and Sheldon, Kruuk & Merilä, 2001 for more extensive treatment of this particular example).

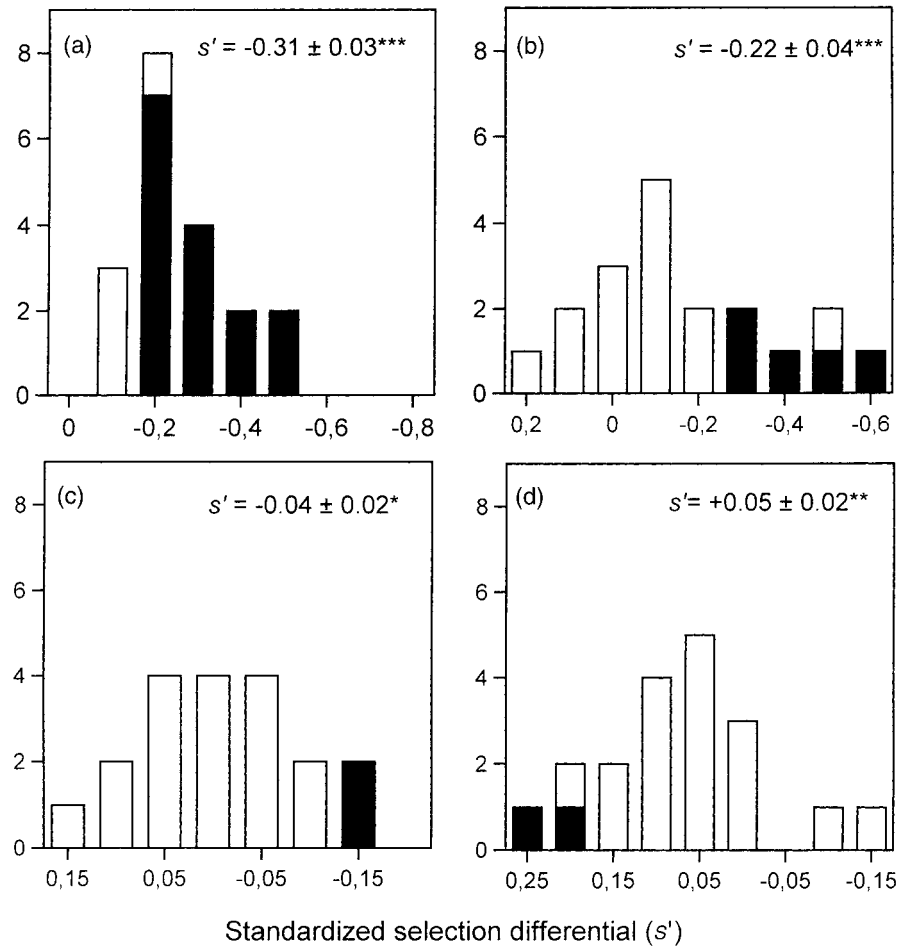

Figure 5. Distribution of annual standardised selection differentials on components of laying date in the collared flycatcher, over 19 years. (a) Selection on phenotypic values via recruitment of offspring, (b) selection on breeding values via recruitment of offspring, (c) selection on phenotypic values via adult survival, (d) selection on breeding values via adult survival. Filled bars show individual selection differentials that are significantly ($p < 0.05$) different from zero. The figures within each panel show the mean standardised selection differential over the 19 year period. $^{*}p < 0.05$, $^{**}p < 0.01$; $^{***}p < 0.001$. Data from Sheldon, Kruuk and Merilä (2001).

Over 20 years (1980–1999) we found no evidence that population mean breeding time had changed with time ($r = -0.18$, $N = 20$, $p = 0.45$). There was a strong negative correlation between the population's mean laying date and the value of the previous winter's North Atlantic Oscillation (NAO) index ($r = -0.58$, $N = 20$, $p = 0.007$), which suggests that the population's behaviour for this character is partly environmentally-determined (see also Przybylo, Sheldon & Merilä, 2000a). There was also no evidence of a change in breeding date over time when the effect of variation in the NAO index was accounted for ($r = -0.24$, $N = 20$, $p = 0.31$). We did find, however, strong selection on laying date when we used the number of offspring recruited to the breeding population as the measure of fitness (Figure 5(a)). This selection was relatively consistent in direction: selection differentials on laying date were statistically significant in 15/19 years, and in all other years they were also negative (Figure 5(a)); the mean standardised selection differential on laying date was -0.306 ± 0.025 SE ($p < 0.0001$). Based on 5889 individual breeding attempts, this represents evidence for strong natural selection at the level of the phenotype (c.f. Kingsolver et al., 2001). Although clutch size and laying date are moderately strongly correlated (both phenotypically and genetically), inclusion of clutch size in multiple regression models has little effect on the estimated strength of selection on laying date: in this case selection

differentials appear to closely approximate selection gradients.

Two previous studies had estimated heritability of laying date in this population; both used traditional parent-offspring regression (in this case mother-daughter). Gustafsson (1986) estimated a heritability of 0.29, while with a larger sample, Merilä and Sheldon (2000) estimated a heritability of 0.41; both estimates are statistically significantly different from zero. We used an animal model approach on this data set to estimate the heritability of laying date: this revealed a heritability of laying date of 0.192 ± 0.036 S.E. ($p < 0.0001$). While lower than the estimates derived from traditional means, this analysis still suggests a substantial additive genetic variance component to breeding time. In combination with the selection gradients reported above, we would expect a response to selection of ≈ 0.059 haldanes: there is no evidence to suggest that such a response has occurred, although the power to reject the null hypothesis is not very large.

However, the application of animal models to the data allows us to test some alternative explanations for microevolutionary stasis. We tested whether there was any evidence for selection acting directly on the breeding values of laying date. Although the mean standardised selection differential on the breeding values for laying date (mean $= -0.218 \pm 0.04$ S.E., Figure 5(b)) was smaller than that on the phenotypic values (paired t-test comparing annual selection differentials: $t = 2.96$, d.f. $= 18$, $p = 0.008$), selection was still consistently negative (Figure 5(b)). Selection differentials on breeding values were significantly negative in 5/19 years, and never significantly greater than zero, and selection differentials on breeding values were positively correlated with those on phenotypic values ($r = 0.75$, $N = 19$, $p = 0.0002$). These data thus suggest that the force of selection on breeding values for laying date is reduced because selection acts more strongly on the environmental deviations, but they still indicate strong natural selection on the underlying additive genetic component of breeding time.

However, the analyses reported above assessed selection via only one component of fitness: that due to production of offspring in each year separately. We asked whether selection acted differently on laying date via adult survival, by calculating selection differentials in each year using relative survival as the measure of fitness. Selection was much weaker, but again we found evidence for an overall pattern of directional selection towards earlier laying dates when we analysed selection on phenotypic values of laying date (Figure 5(c)). However, the pattern was different when selection on breeding values for laying date via this fitness pathway was examined (Figure 5(d)). In this case, the mean standardised selection gradient was positive, implying that individuals with a positive breeding value for laying date (i.e., a breeding value causing them to breed later than the population mean, on average) were more likely to survive to the next breeding season. This effect will tend to counteract the negative selection differential on breeding values due to recruitment of offspring. The extent to which the two types of selection balance each other will depend upon the age-structure and age-specific fecundity of the population; this question is addressed further in Sheldon, Kruuk and Merilä (2001).

To sum up, in this case analysis of selection on breeding values was informative, since it suggested that explanations for microevolutionary stasis of breeding time based on selection acting on environmental deviations (e.g., Price, Kirkpatrick & Arnold, 1988) do not offer a complete explanation. An additional genetic constraint was uncovered: selection on breeding values acts in different directions depending on which component of fitness is measured. Individuals with large negative breeding values for laying date (early breeders) recruit more offspring per breeding attempt, but have shorter lifespans; individuals with large positive breeding values for laying date recruit fewer offspring per breeding attempt, but live for longer. These patterns imply the existence of a negative genetic correlation between current and future reproductive output. Interestingly, this pattern would not have been detected had analyses been restricted to phenotypes alone, since selection on laying date phenotypes was negative for both fitness pathways. It has been recognised for some time that studying phenotypes can give a misleading impression about life-history evolution, if there is appreciable variation in resource acquisition relative to resource allocation (van Noordwijk & de Jong, 1986). Our data are consistent with the suggestion that selection on laying date at the level of the phenotype largely represents selection due to variation in resource acquisition.

4. *Concealed evolution in flycatchers?*

Our final example of the utility of animal model approaches for studies of real-time evolution in the wild comes from our recent work attempting to understand the apparent lack of long-term change

in the predicted direction of mean condition index (relative body mass) in collared flycatcher fledglings. As shown above, there is strong directional selection acting on the condition index, both on phenotypic (Figure 3(b)) and breeding (Figure 3(d)) values, and the trait is heritable (Merilä, Kruuk & Sheldon, 2001a; Table 1). However, unexpectedly, the mean condition index has declined over the course of the study period (Figure 6(a)). Hence, the situation is parallel to that observed in the case of body size of snow geese (Cooch et al., 1991) and barnacle geese (Larsson et al., 1998). In order to test the conjecture that evolution at the genotypic level had actually taken place, but had been masked by changes in environment during the same period, we estimated the mean breeding values for nestlings born in different years and plotted them against time. The mean breeding value of the condition index is a positive function of time (Figure 6(b)), supporting this suggestion. Thus in this case it appears that microevolution has occurred in the expected direction, although this is not detectable (Merilä, Kruuk & Sheldon, 2001b). One possible explanation for this is that some aspect of the environment affecting offspring condition has deteriorated over the course of study period, effectively masking the evolution seen in offspring breeding values Merilä, Kruuk & Sheldon, 2001b). We cannot currently identify cause causal factor behind the apparent environmental deterioration (which is reflected e.g., in a linear reduction of breeding success over time (Merilä, Kruuk & Sheldon, 2001b)), but we suspect that this might relate to decreased food availability due to increased degree of asynchrony between oak bud burst dates and caterpillar (the main food supply of nestling flycatchers) emergence dates (Visser & Holleman, 2001). Whatever the proximate cause, this example illustrates the potential of using estimates of breeding values for detecting 'concealed' microevolutionary changes in wild populations. Again, we can rule out explanations 1–3 in this case, and we have also data (Merilä, Kruuk & Sheldon, 2001b) to indicate that negative genetic correlations with other fitness components are unlikely explanation for lack of observed selection response (explanation 4).

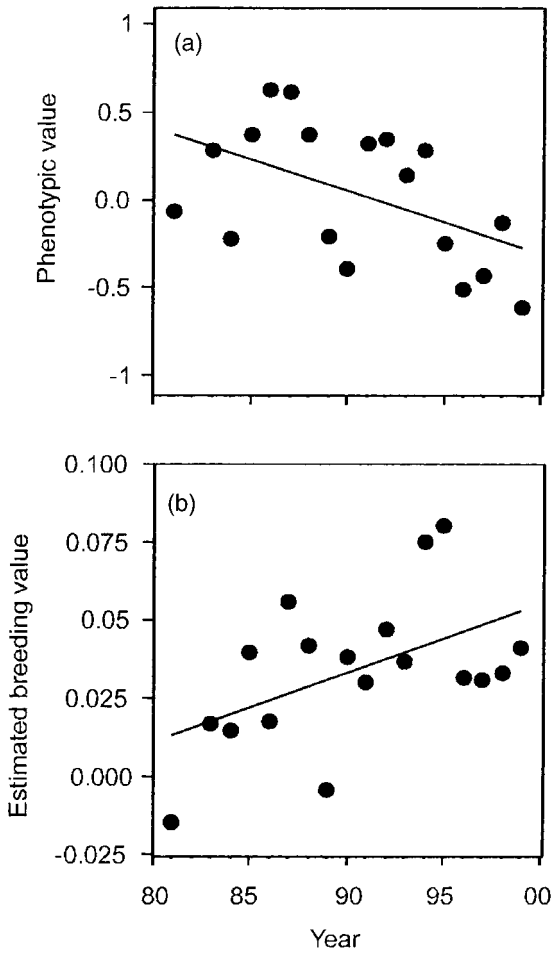

Figure 6. Mean collared flycatcher nestling condition as a function of time. (a) Mean phenotypic value (linear regression: $b = -0.036 \pm 0.015$ S.E., $F_{1,17} = 5.32$, $p = 0.032$), (b) mean estimated breeding value (EBV; linear regression: $b = 0.0022 \pm 0.0009$, $F_{1,17} = 5.65$, $p = 0.030$). Adapted from Merilä, Kruuk and Sheldon (2001b).

Future directions

In the examples outlined above, we have shown that methods adopted from animal breeding sciences offer the prospect of progress in testing alternative explanations for apparent microevolutionary stasis in wild populations. Although these methods are unlikely to solve all of the problems encountered, they do allow explicit tests of some critical predictions which have previously only existed as verbal (e.g., Alatalo, Gustafsson & Lundberg, 1990) or analytical arguments (e.g., Price, Kirkpatrick & Arnold, 1988). In three out of four of the cases examined, we were able to reject the hypothesis that stasis is due entirely to selection on environmental deviations; whereas in one case this hypothesis was supported. However, it is worth noting that in all cases, selection on phenotypic values was stronger than on breeding values, implying that the path structure (explanation 3) suggested by Price, Kirkpatrick and Arnold (1988) has some validity.

The results further show that the concern about correct estimation of trait heritabilities (explanation 1) is an issue to be taken seriously: in all cases where animal model analyses have been applied (Milner et al., 2000; Kruuk et al., 2000, 2001; Coltman et al., 2001; Kruuk, Merilä & Sheldon, 2001; Merilä, Kruuk & Sheldon, 2001; Sheldon, Kruuk & Merilä, 2001), a significant proportion of the phenotypic variance has been attributable to common environment and maternal effects. Our examples about collared flycatcher tarsus length and laying dates are cases in point – in both instances, the heritabilities returned by animal model analyses were substantially lower than the earlier estimates derived using conventional methods (e.g., Merilä & Sheldon, 2000). Nevertheless, heritabilities were still substantial, and the lack (or biased estimation) of genetic variability is unlikely to be a general explanation for the microevolutionary stasis observed in the long-term studies (Table 1). Nevetheless, the role of more complex maternal effect influences known to cause unexpected evolutionary dynamics remain to be investigated.

A striking pattern emerging from the Table 1 is that many heritable and/or apparently heritable traits under consistent directional selection do not only show microevolutionary stasis, but they have actually changed in directions opposite to that predicted on the basis of Equation (1) alone. Although these changes have been often explained in terms of plastic phenotypic responses to changed environmental conditions (e.g., Cooch et al., 1991; Larsson et al., 1998), our example concerning directional change in the mean breeding value of condition in the collared flycatcher population over time suggests that it would be premature to take them to indicate that no evolutionary transformations have taken place. In other words, despite the fact that changes at the phenotypic level have occurred in one direction, this does not preclude the possibility that changes at the genotypic level have occurred in the other direction. Hence, as argued by Cooke et al. (1990) in the context of clutch size changes in the snow goose, changes in environmental conditions over time may act to conceal microevolution. Although there is nothing new in this perspective if we consider the view that organisms or populations are likely to be constantly evolving just to stay adapted to the current environment (Lewontin, 1978), our treatment is the first one to attempt to test this with an animal model approach and estimated breeding values. Clearly, although more work is needed to elucidate how common this type of 'concealed' evolution might be in the long-term data sets, our example serves to indicate that explanation 5 may account for some of the stasis observed in the Table 1.

We hope that this paper also makes it clear that the value of long-term studies is likely to increase with time not only due to the amount of accumulated data and associated increase in statistical power, but also because of the invention of new uses for the data (see also Perrins, 1994). In particular, we have illustrated how two long-term data sets have provided the basis for studies of natural populations which would not be possible with data collected over shorter time frames. The reasons for this are two fold. First, the estimation of quantitative genetic parameters, and breeding values in particular, with the animal model approach is most effectively done with long pedigrees and unless such are available, the advantage over traditional methods is unlikely to be large (Knott et al., 1995). The other reason is that only long-term data sets provide time series long enough to be able to test for temporal variation in selection pressures, and to have a reasonable likelihood of detecting any response to selection. To this end, we believe that further analysis of already existing long-term data sets, such as those included in Clutton-Brock (1988), using the methods described above would be a significant and desirable contribution to the otherwise small literature (c.f. Table 1) relevant to problems outlined in this paper.

Finally, it is perhaps worth emphasising the fact that the analyses of the effects of natural selection on genetic and environmental components of variance presented above are, to our knowledge, the only direct empirical evidence from wild populations that natural selection is reducing the additive genetic variance in a trait. This suggests an additional way in which long-term studies of wild populations can contribute to our understanding of the interplay between natural selection and genetic variation, namely, in investigating how selection moulds the genetic architecture of different types of traits (c.f. Merilä & Sheldon, 1999). This type of information is valuable not only due its fundamental interest, but also from the point of view of predicting how the long-term dynamics of genetic variability in different traits might change as a function of time, and contribute to long-term persistence and adaptation of populations under changing environmental conditions (e.g., Bürger & Lynch, 1998). The animal model approach applied to long-term data sets is not a panacea that is likely to solve all problems faced by the studies of contemporary

microevolution. Nevertheless, in combination with indirect approaches (Table 1) and carefully planned experimentation in the wild, it can provide insights into processes and factors central to our understanding of microevolution which would be not obtainable by any other means, or which lack some of the biological realism when organisms (and genes) are studied away from their natural environments.

Conclusions

We have identified a number of studies in which the data seem to conflict with theoretical expectations about the direction and rate of evolution in the traits in question. Using these observations as a starting point, we have outlined a general framework for attempting to understand why real-time microevolution (*sensu stricto*, note that there is good evidence for microevolutionary transformations from other sorts of approaches) is so rarely observed in the wild, with the emphasis on the application of statistical methods from animal breeding sciences to long-term data sets accumulated from studies of individually marked vertebrate populations. Application of these methods to two long-term data sets suggests that, although applicable in some cases, certain previously-advocated hypotheses are unlikely to provide any general explanation for the lack of observed microevolution in these studies. However, although more research is required to reach any general conclusions, we suggest that changes occurring at the genetic level despite the appearance of stasis (or even opposite changes) at the phenotypic level may be part of the answer. Nevertheless, the general paucity of direct evidence for microevolutionary changes in long-term data sets stands in contrast to overwhelming evidence for ubiquitous natural selection and adaptation in wild (e.g., Bone & Farres, 2001; Kinnison & Hendry, 2001; Reznick & Ghalamber, 2001), and calls for further investigations as to why evolutionary transformations in nature should be so difficult to observe.

Acknowledgements

Our research has been supported by the grants from the Swedish Natural Science Research Council, the UK Natural Environment Research Council and the Royal Society of London. LEBK and BCS are Royal Society University Research Fellows. We thank Sue Brotherstone and Peter Visscher for statistical advice, Andrew Hendry and an anonymous referee for suggestions that improved the manuscript.

References

Alatalo, R.V., L. Gustafsson & A. Lundberg, 1990. Phenotypic selection on heritable size traits. Environmental variance and genetic response. Am. Nat. 135: 464–471.

Andersson, S., P. Waldmann & H.C. Pretince, 2000. The structure of quantitative genetic variance within the narrow endemic *Silene diclinis* (Caryophyllaceae). (submitted).

Arnold, S.J., 1981. Behavioral variation in natural populations. II: the inheritance of a feeding response in crosses between geographic races of the garter snake, *Thamnophis elegans*. Evolution 35: 510–515.

Arnold, S.J. & M.J. Wade, 1984a. On the measurement of natural and sexual selection: theory. Evolution 38: 709–719.

Arnold, S.J. & M.J. Wade, 1984b. On the measurement of natural and sexual selection: applications. Evolution 38: 720–734.

Barton, N.H. & K.S. Gale, 1993. Genetic analysis of hybrid zones, pp. 13–45 in Hybrid Zones and the Evolutionary Process, edited by R.G. Harrison. Oxford Univ. Press, Oxford.

Barton, N.H. & L. Partridge, 2000. Limits to natural selection. Bioessays 22: 1075–1084.

Beebee, T.J.C., 1995. Amphibian breeding and climate. Nature 374: 219–220.

Bernardo, J., 1996. Maternal effects in animal ecology. Amer. Zool. 36: 83–105.

Berthold, P., A.J. Helbig, G. Mohr & U. Querner, 1992. Rapid microevolution of migratory behavior in a wild bird species. Nature 360: 668–670.

Berven, K.A. & D.E. Gill, 1983. Interpreting geographic variation in life history traits. Am. Zool. 23: 85–97.

Blair, H.T. & E.J. Pollak, 1984. Estimation of genetic trend in a selected population with and without the use of a control population. J. Anim. Sci. 58: 878–886.

Blondel, J., P.C. Dias, P. Perret, M. Maistre & M.M. Lambrechts, 1999. Selection-based biodiversity at a small spatial scale in a low-dispersing insular bird. Science 285: 1399–1402.

Bone, E. & A. Farres, 2001. Trends and rates of microevolution in plants. Genetica 112–113: 165–182.

Brotherstone, S., C.M. McManus & W.B. Hill, 1990. Estimation of genetic parameters for linear and miscellaneous type traits in Holstein–Friesian dairy cattle. Livestock Prod. Sci. 26: 177–192.

Brown J.L., S.H. Li & N. Bhagabati, 1999. Long-term trend toward earlier breeding in an American bird: a response to global warming? Proc. Natl. Acad. Sci. U.S.A. 96: 5565–5569.

Boyce, M.S. & C.M. Perrins, 1987. Optimizing great tit clutch size in a fluctuating environment. Ecology 68: 142–153.

Butlin, R.K. & T. Tregenza, 1998. Levels of genetic polymorphism: marker loci versus quantitative traits. Phil. Trans. R. Soc. Lond. B. 353: 187–198.

Bürger, R. & M. Lynch, 1997. Adaptation and extinction in changing environments, pp. 209–240 in Environmental Stress, Adaptation and Evolution, edited by R. Bijlsma & V. Loeschcke. Springer Verlag, Berlin.

Caroll, S.P., H. Dingle, T.R. Famula & C.W. Fox, 2001. Genetic architecture of adaptive differentiation in evolving host races

of the soapberry bug, *Jadera haematoloma*. Genetica 112–113: 257–272.
Clutton-Brock, T.H., 1988. Reproductive Success. Chicago Univ. Press, Chicago.
Clutton-Brock, T.H., M. Major, S.D. Albon & F.E. Guinness, 1987. Early development and population dynamics in red deer. I: density-dependent effects on juvenile survival. J. Anim. Ecol. 56: 53–67.
Clutton-Brock, T.H. & S.D. Albon, 1989. Red Deer in the Highlands. Blackwell Scientific Publications, Oxford.
Coltman, D.W., J. Pilkington, L.E.B. Kruuk, K. Wilson & J.M. Pemberton, 2001. Positive genetic correlations between parasite resistance and body size in a free-living ungulate population. Evolution (in press).
Conover, D.O. & E.T. Schultz, 1995. Phenotypic similarity and the evolutionary significance of countergradient variation. Trends Ecol. Evol. 10: 248–252.
Cooch, E.G., D.B. Lank, R.F. Rockwell & F. Cooke, 1991. Long-term decline in body size in a snow goose population: evidence of environmental degradation? J. Anim. Ecol. 60: 483–496.
Cooke, F. & C.S. Findlay, 1982. Polygenic variation and stabilizing selection in a wild population of lesser snow geese (*Anser caerulescens caerulescens*). Am. Nat. 120: 543–547.
Cooke, F., R.F. Rockwell & D.B. Lank, 1995. The Snow Geese of La Pérouse Bay. Natural Selection in the Wild. Oxford Univ. Press, Oxford.
Cooke, F., P.D. Taylor, C.M. Frances & R.F. Rockwell, 1990. Directional selection and clutch size in birds. Am. Nat. 136: 261–267.
Crespi, B.J., 1990. Measuring the effect of natural selection on phenotypic interaction systems. Am. Nat. 135: 32–47.
Daan, S., C. Dijkstra & J.M. Tinbergen, 1990. Family planning in the kestrel (*Falco tinnunculus*): the ultimate control of covariation of laying date and clutch size. Behaviour 114: 83–116.
Dhondt, A.A., F. Adriansen, E. Matthysen & B. Kempenaers, 1990. Nonadaptive clutch sizes in tits. Nature 348: 723–725.
Dhondt, A.A., R. Eyckerman & J. Huble, 1979. Will great tits become little tits? Biol. J. Linn. Soc. 11: 289–294.
Dias, P.C. & J. Blondel, 1996. Local specialization and maladaptation in the Mediterranean blue tit (*Parus caeruleus*). Oecologia 107: 79–86.
De Meester, L., 1996. Local genetic differentiation and adaptation in freshwater zooplankton populations: patterns and processes. Écoscience. 3: 385–399.
Ebert, D., 1995. Virulence and local adaptation of a horizontally transmitted parasite. Science 265: 1084–1086.
Edwards, S.V. & M. Kot, 1995. Comparative methods at the species level: geographic variation in morphology and group size in grey-crowned babblers (*Pomatostomus temporalis*). Evolution 49: 1134–1146.
Endler, J.A., 1977. Geographic Variation, Speciation, and Clines. Princeton Univ. Press, Princeton.
Endler, J.A., 1980. Natural selection on color patterns in *Poecilia reticulata.* Evolution 34: 76–91.
Endler, J.A., 1986. Natural Selection in Wild. Princeton Univ. Press, Princeton.
Falconer, D.S. & T. Mackay, 1996. Introduction to Quantitative Genetics. Longman, N.Y.
Ferraz, J.B.S. & Johnson, R.K., 1993. Animal model estimation of genetic parameters and response to selection for litter size and weight, growth, and backfat in closed seedstock populations of large white and landrace swine. J. Anim. Sci. 71: 850–858.
Findlay, C.S. & F. Cooke, 1982. Breeding synchrony in the lesser snow goose (*Anser caerulescens caerulescens*). I: genetic and environmental components of hatch date variability and their effects on hatch date synchrony. Evolution 36: 342–351.
Fisher, R.A., 1958. The Genetical Theory if Natural Selection. Clarendon, London.
Flux, J.E.C. & M.M. Flux, 1982. Artificial selection and gene flow in wild starlings, *Sturnus vulgaris*. Naturwissenschaften 69: 96–97.
Forchhammer, M.C., E. Post & N.C. Stenseth, 1998. Breeding phenology and climate... . Nature 391: 29–30.
García-Ramos, G. & M. Kirkpatrick, 1997. Genetic models of adaptation and gene flow in peripheral populations. Evolution 51: 21–28.
Gibbs, H.L., 1988. Heritability and selection on clutch size in Darwin's medium ground finches (*Geospiza fortis*). Evolution 42: 750–762.
Gilchrist, A.S. & L. Partridge, 1999. A comparison of the genetic basis of wing size divergence in three parallel body size clines of *Drospohila melanogaster*. Genetics 153: 1775–1787.
Grant, B.R. & P.R. Grant, 1993. Evolution of Darwin's finches caused by a rare climatic event. Proc. R. Soc. Lond. B 251: 111–117.
Grant, P.R. & B.R. Grant, 1995. Predicting microevolutionary responses to directional selection on heritable variation. Evolution 49: 241–251.
Groeneveld, E., 1995. REML VCE, a Multivariate Multi-Model Restricted Maximum Likelihood (Co)Variance Component Estimation Package, Version 3.2 User's Guide. http://www.tzv.fal.de/institut/genetik/pub/eg/vce/manual/manual.html. Mariensee, Germany: Institute of Animal Husbandry and Animal Behaviour, Federal Research Center of Agriculture (FAL).
Groeneveld, E. & M. Kovac, 1990. A generalized computing procedure for setting up and solving mixed linear-models. J. Dairy Sci. 73: 513–531.
Gustafsson, L., 1986. Lifetime reproductive success and heritability: empirical support for Fisher's fundamental theorem. Am. Nat. 128: 761–764.
Hairston N.G. & W.E. Walton, 1986. Rapid evolution of a life history trait. Proc. Natl. Acad. Sci. U.S.A. 83: 4831–4833.
Hedrick, P.W., 1999. Highly variable loci and their interpretation in evolution and conservation. Evolution 53: 313–318.
Hendry, A.P. & M.T. Kinnison, 1999. Perspective: the pace of modern life: measuring rates of contemporary microevolution. Evolution 53: 1637–1653.
Hendry, A.P., T. Troy & E.B. Taylor, 2001. Population mixing and the adaptive divergence of quantitative characters in discrete populations: a theoretical framework for empirical tests. Evolution 55: 459–466.
Hill, W.G. & A. Caballero, 1992. Artificial selection experiments. Annu. Rev. Ecol. Syst. 23: 287–310.
Huey, R.B., G.W. Gilchrist, M.L. Carlson, D. Berrigan & L. Serra, 2000. Rapid evolution of a geographic cline in size in an introduced fly. Science 287: 308–309.
Hughes, L., 2000. Biological consequences of global warming: is the signal already apparent? Trends Ecol. Evol. 15: 56–61.
Houle, D., 1991. What genetic correlations are made of, and why it matters. Evolution 45: 630–648.
James, F.C., 1983. Environmental component of morphological differentiation in birds. Science 221: 184–186.
Kinnison M.T. & A.P. Hendry, 2001. The pace of modern life. II: from rates of contemporary microevolution to pattern and process. Genetica 112–113: 145–164.

Kingsolver, J.G., H.E. Hoekstra, J.M. Hoekstra, D. Berrigan, S.N. Vignieri, C.E. Hill, A. Hoang, P. Gibert & P. Beerli, 2001. The strength of phenotypic selection in natural populations. Am. Nat. 157: 245–261.

Kirkpatrick, M. & R. Lande, 1989. The evolution of maternal characters. Evolution 43: 485–503.

Klomp, H., 1970. The determination of clutch size in birds: a review. Ardea 58: 1–125.

Knott, S.A., R.M. Sibly, R.H. Smith & H. Møller, 1995. Maximum likelihood estimation of genetic parameters in life-history studies using the 'Animal Model'. Funct. Ecol. 9: 122–126.

Kruuk, L.E.B., T.H. Clutton-Brock, J. Slate, J.M. Pemberton, S. Brotherstone & F.E. Guinness, 2000. Heritability of fitness in a wild mammal population. Proc. Natl. Acad. Sci. U.S.A. 97: 698–703.

Kruuk, L.E.B., J. Merilä & B.C. Sheldon, 2001. Phenotypic selection on heritable size traits revisited. Am. Nat. (in press).

Kruuk, L.E.B., J.M. Pemberton, J. Slate & T.H. Clutton-Brock, 2001. Antler size in red deer: heritability and selection without evolution. (Submitted).

Lack, D., 1968. Ecological Adaptations for Breeding in Birds. Methuen, London.

Lande, R., 1980. Genetic variation and phenotypic evolution during allopatric speciation. Am. Nat. 116: 463–479.

Lande, R. & S.J. Arnold, 1983. The measurement of selection on correlated characters. Evolution 37: 1210–1226.

Larsson, K., 1993. Inheritance of body size in the barnacle goose under different environmental conditions. J. Evol. Biol. 6: 195–208.

Larsson, K., H.P. van der Jeugd, I.T. van der Veen & P. Forslund, 1998. Body size declines despite positive directional selection on heritable size traits in a barnacle goose population. Evolution 52: 1169–1184.

Lewontin, R.C., 1978. Adaptation. Sci. Am. 239: 212–230.

Linhart, Y.B. & M.C. Grant, 1996. Evolutionary significance of local genetic differentiation in plants. Annu. Rev. Ecol. Syst. 27: 237–277.

Lively, C.M. & J. Jokela, 1996. Clinal variation for local adaptation in a host-parasite interaction. Proc. R. Soc. Lond. B 263: 891–897.

Losos, J.B., K.I. Warheit & T.W. Schoener, 1997. Adaptive differentiation following experimental island colonization in *Anolis* lizards. Nature 387: 70–73.

Losos, J.B., T.R. Jackman, A. Larson, K. de Queiroz & L. Rodríguez-Schettino, 1998. Contingency and determinism in replicated adaptive radiations of island lizards. Science 279: 2115–2118.

Losos, J.B., D.A. Creer, D. Glossip, R. Goellner, A. Hampton, G. Roberts, N. Haskell, P. Taylor & J. Ettling, 2000. Evolutionary implications of phenotypic plasticity in the hindlimb of the lizard *Anolis sagrei*. Evolution 54: 301–305.

Lynch, M., 1984. The selective value of alleles underlying polygenic traits. Genetics 108: 1021–1033.

Lynch, M. & B. Walsh, 1998. Genetics and Analysis of Quantitative Traits. Sinauer Association, Sunderland, Mass.

Lynch, M., M. Pfrender, K. Spitze, N. Lehman, J. Hicks, D. Allen, L. Latta, M. Ottene, F. Bogue & J. Colbourne, 1999. The quantitative and molecular genetic architecture of a subdivided species. Evolution 53: 100–110.

Martinez, V., L. Bünger, L. & W.G. Hill, 2000. Analysis of response to 20 generations of selection for body composition in mice: fit to infinitesimal model assumptions. Genet. Sel. Evol. 32: 3–21.

McCleery, R.H. & C.M. Perrins, 1998. ...temperature and egg-laying trends. Nature 391: 30–31.

McKenzie, J.A. & P. Batterham, 1994. The genetic, molecular and phenotypic consequences of selection for insecticide resistance. Trends Ecol. Evol. 9: 166–169.

Merilä, J., 1997. Quantitative trait and allozyme divergence in the greenfinch (*Carduelis chloris*, Aves: Fringillidae). Biol. J. Linn. Soc. 61: 243–266.

Merilä, J. & M. Björklund, 1999. Population divergence and morphometric integration in the Greenfinch – evolution against the trajectory of least resistance? J. Evol. Biol. 12: 103–112.

Merilä, J., M. Björklund & L. Gustafsson, 1994. Evolution of morphological differences with moderate genetic correlations among traits as exemplified by flycatchers (Aves: Muscicapidae). Biol. J. Linn. Soc. 52: 19–30.

Merilä, J. & B.C. Sheldon, 1999. Genetic architecture of fitness and non-fitness traits – empirical patterns and development of ideas. Heredity 83: 103–109.

Merilä, J. & B.C. Sheldon, 2000. Lifetime reproductive success and heritability in nature. Am. Nat. 155: 301–310.

Merilä, J. & B.C. Sheldon, 2001. Avian quantitative genetics. Curr. Ornithol. 9: 179–255.

Merilä, J., B.C. Sheldon & H. Ellegren, 1998. Quantitative genetics of sexual size dimorphism in the collared flycatcher, *Ficedula albicollis*. Evolution 52: 870–876.

Merilä, J., L.E.B. Kruuk & B.C. Sheldon, 2001a. Natural selection on the genetical component of variance in body condition in a wild bird population. J. Evol. Biol. (in press).

Merilä, J., L.E.B. Kruuk & B.C. Sheldon, 2001b. Cryptic evolution in a wild bird population. Nature 412: 76–79.

Meyer, K., 1991. Estimating variances and covariances for multivariate animal models by restricted maximum likelihood. Genet. Selec. Evol. 23: 67–83.

Meyer, K., 1997. DFREML: programs to estimate variance components by restricted maximum likelihood using a derivative-free algorithm. User notes, version 3.0α. http:// agbu.une.edu/ kmeyer/dfreml.html.

Meyer, K. & W.G. Hill, 1991. Mixed model analysis of a selection experiment for food intake in mice. Genet. Res. 57: 71–81.

Milkman, R., 1982. Toward unified selection theory, pp. 105–118 in Perspectives on Evolution, edited by R. Milkman. Sinaeur Association, Sunderland, Mass.

Milner, J.M., S.D. Albon, A.W. Illius, J.M. Pemberton & T.H. Clutton-Brock, 1999. Repeated selection of morphometric traits in the Soay sheep on St. Kilda. J. Anim. Ecol. 68: 472–488.

Milner, J.M., S. Brotherstone, J.M. Pemberton & S.D. Albon, 2000. Estimating variance components and heritabilities in the wild: a case study using the 'animal model' approach. J. Evol. Biol. 13: 804–813.

Mousseau, T.A. & C.W. Fox, 1998. Maternal effects as adaptations. Oxford Univ. Press, Oxford.

Norris, K.J., 1993. Seasonal variation in the reproductive success of the blue tits: an experimental study. J. Anim. Ecol. 62: 287–294.

Ollivier, L., 1999. On the use of animal models in the analysis of selection experiments. Genet. Sel. Evol. 31: 135–148.

Partridge, L. & J.A. Coyne, 1997. Bergmann's rule in ectotherms: is it adaptive? Evolution. 51: 632–635.

Perrins, C.M., 1994. Tits and their caterpillar food supply. Ibis 133 (suppl. 1): 49–54.

Price, D.K., 1996. Sexual selection, selection load and quantitative genetics of zebra finch bill colour. Proc. Roy. Soc. Lond. B 263: 217–221.

Price, T.D., P.R. Grant, H.L. Gibbs & P.T. Boag, 1984. Recurrent patterns of natural selection in a population of Darwin's finches. Nature 309: 787–789.

Price, T.D., M. Kirkpatrick & S.J. Arnold, 1988. Directional selection and the evolution of breeding date in birds. Science 240: 798–799.

Price, T.D. & L. Liou, 1989. Selection on clutch size in birds. Am. Nat. 134: 950–959.

Price, T.D. & T. Langen, 1992. Evolution of correlated characters. Trends Ecol. Evol. 7: 307–310.

Przybylo, R., B.C. Sheldon & J. Merilä, 2000a. Climatic influences on breeding and morphology : evidence for phenotypic plasticity. J. Anim. Ecol. 69: 395–403.

Przybylo, R., B.C. Sheldon & J. Merilä, 2000b. Patterns of natural selection on body size of male and female collared flycatchers – a long term study. Biol. J. Linn. Soc. 69: 213–232.

Rausher, M.D., 1992. The measurement of selection on quantitative traits: biases due to the environmental covariances between traits and fitness. Evolution 46: 616–626.

Réale, D., M. Festa-Bianchet & J.T. Jorgenson, 1999. Heritability of body mass varies with age and season in wild bighorn sheep. Heredity 83: 526–532.

Reznick, D.N. & C.K. Ghalambor, 2001. The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. Genetica 112–113: 183–198.

Reznick, D.N., F.H. Shaw, F.H. Rodd & R.G. Shaw, 1997. Evaluation of the rate of evolution in a natural population of guppies. Science 275: 1934–1937.

Riechert, S.E. 1993. Investigation of potential gene flow limitation of behavioural adaptation in an aridlands spider. Behav. Ecol. Sociobiol. 32: 355–363.

Roff, D.A., 1997. Evolutionary Quantitative Genetics. Chapman & Hall, N.Y.

Roff, D.A., 2000. The evolution of the G-matrix: selection or drift? Heredity 84: 135–142.

Rowe, L., D. Ludwig & D. Schluter, 1994. Time, condition, and the seasonal decline of avian clutch size. Am. Nat. 143: 698–722.

Schluter, D., 1996. Adaptive radiation along genetic lines of least resistance. Evolution. 50: 1766–1774.

Schluter, D., 2000. The Ecology of Adaptive Radiation. Oxford Univ. Press, Oxford.

Schluter, D., T.D. Price & L. Rowe, 1991. Conflicting selection pressures and life-history trade-offs. Proc. R. Soc. Lond. B 246: 11–17.

Schultz, E.T., 1993. The effect of birth date on fitness of female dwarf perch, *Micrometrus minimus* (Perciformis: Embiotocidae). Evolution 47: 520–539.

Semlitsch, R.D. & H.M., Wilbur, 1989. Artificial selection for paedomorphosis in the salamander *Ambystoma talpoideum*. Evolution 43: 105–112.

Shaw, R.G. & D.L. Byers, 1998. Genetics of maternal and paternal effects, pp. 97–111 in Maternal Effects as Adaptations, edited by T.A. Mousseau & C.W. Fox. Oxford Univ. Press, Oxford.

Sheldon, B.C., L.E.B. Kruuk & J. Merilä, 2001. Natural selection and the evolution of avian breeding time in a variable environment. (submitted).

Sinervo, B. & P. Doughty, 1996. Interactive effects of offspring size and timing of reproduction on offspring reproduction: experimental, maternal and quantitative genetic aspects. Evolution 50: 1314–1327.

Sorenson, D.A. & B.W. Kennedy, 1984. Estimation of genetic variances from unselected and selected populations. J. Anim. Sci. 58: 1097–1106.

Southwood, O.I. & B.W. Kennedy, 1991. Genetic and environmental trends for litter size in swine. J. Anim. Sci. 69: 3177–3182.

Spitze, K., 1993. Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation. Genetics 135: 367–374.

Storfer, A. & A. Sih, 1998. Gene flow and ineffective antipredator behavior in a stream-breeding salamander. Evolution 52: 558–565.

Trussell, G.C. & Smith, L.D., 2000. Induced defenses in response to an invading crab predator: an explanation of historical and geographic phenotypic change. Proc. Natl. Acad. Sci. U.S.A. 97: 2123–2127.

van der Werf, J.H.J. & I.J.M. de Boer, 1990. Estimation of additive genetic variance when base populations are selected. J. Anim. Sci. 68: 3124–3132.

van Noordwijk, A.J., 1990. The methods of genetical ecology applied to the study of evolutionary change, pp. 291–319 in Population Biology, Ecological and Evolutionary View Points, edited by K. Wöhrman & S.K. Jain. Springer-Verlag, Berlin, Heidelberg.

van Noordwijk, A.J., J.H. van Balen & W. Scharloo, 1981. Genetic variation in the timing of reproduction in the great tit. Oecologia 49: 158–166.

van Noordwijk, A.J. & G. de Jong, 1986. Acquisition and allocation of resources: their influence on variation in life-history tactics. Am. Nat. 128: 137–142.

van Tienderen, P.H. & G. de Jong, 1994. A general model of the relation between phenotypic selection and genetic response. J. Evol. Biol. 7: 1–12.

Verhulst, S., J.H. van Balen & J.M. Tinbergen, 1995. Seasonal decline in reproductive success in the great tit: variation in time or quality? Ecology 76: 2392–2403.

Wade, M.J., 1998. The evolutionary genetics of maternal effects, pp. 5–21 in Maternal Effects as Adaptations, edited by T.A. Mousseau & C.W. Fox. Oxford Univ. Press, Oxford.

Waldmann, P. & S. Andersson, 1998. Comparison of quantitative genetic variation and allozyme diversity within and between populations of *Scabiosa canescens* and S. *columbaria*. Heredity 81: 79–86.

Ward, R.D., D.O.F. Skibinsky & M. Woodwark, 1992. Protein heterozygosity, protein structure and taxonomic differentiation. Evol. Biol. 26: 3–131.

Wiggins, D.A., T. Pärt & L. Gustafsson, 1994. Seasonal decline in collared flycatcher *Ficedula albicollis* reproductive success: an experimental approach. Oikos 70: 359–364.

Wikelski, M. & C. Thom, 2000. Marine iguanas shrink to survive El Nino. Nature 403: 37–38.

Visser, M.E. & L.J.M., Holleman, 2001. Warmer springs disrupt the synchrony of oak and winter moth phenology. Proc. R. Soc. Lond. B. 268: 289–294.